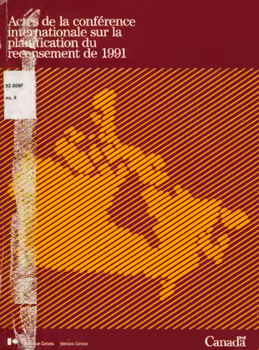


Actes de la conférence internationale sur la planification du recensement de 1991

Statistique Canada Statistics Canada

Canada



Actes de la conférence internationale sur la planification du recensement de 1991

Ottawa, 8-11 octobre 1985

Publication autorisée par le ministre des Approvisionnements et Services Canada

Avril 1987

ISBN 0-662-93889-5

Ottawa

This publication is available in English upon request.

Message de M. Edward T. Pryor

La Conférence internationale sur la planification du recensement de 1991, tenue en octobre 1985, a marqué pour Statistique Canada le point de départ de la planification et de l'élaboration des prochains recensements décennaux de la population et de l'agriculture, qui auront lieu au mois de juin 1991.

Le recensement de la population est la méthode de collecte de données la plus importante et la plus visible de Statistique Canada. Il constitue la pierre angulaire sur laquelle repose l'ensemble du programme de statistiques sociales et fournit des données à des milliers d'utilisateurs de tous les coins du pays. Le recensement de l'agriculture joue un rôle semblable en prenant le pouls de l'industrie agricole au Canada. La conférence internationale représentait la première tentative visant à définir la forme que prendra le recensement de 1991 et à déterminer la meilleure façon d'aborder les nombreuses questions que nous devons régler relativement aux aspects techniques, au contenu et aux politiques.

La participation de spécialistes du recensement d'autres pays, comme le Royaume-Uni, les États-Unis, l'Australie et la Suède, a grandement aidé Statistique Canada. Non seulement ces experts nous ont apporté leur expérience considérable, de nouvelles idées et de nouvelles méthodes, mais ils ont aussi discuté avec nous des problèmes communs qui nous préoccupent tous. De tels partages de renseignements et d'expérience sont très profitables à tous les pays participants, et continueront de l'être.

Je voudrais remercier les participants et les organisateurs qui ont fait un succès de la conférence. Grâce à leur travail, nous avons établi des bases solides pour 1991.

Edward T. Pryor
Directeur général
Statistique démographique et recensement

TABLE DES MATIÈRES

Page

TABLE DES MATIÈRES - suite

TABLE DES MATIÈRES - fin

Page

Introduction

Ce manuel se veut ún compte rendu de la conférence sur la planification du recensement de 1991 qui a eu lieu, du 8 au 11 octobre 1985, à l'intérieur des bureaux de Statistique Canada à Ottawa. Ce manuel a été assemblé de façon à respecter le déroulement de la conférence telle qu'elle fut présentée en 1985.

Cette conférence avait pour but d'aider Statistique Canada à entreprendre la planification du recensement de 1991 tout en poursuivant les activités reliées au recensement de 1986.

Tout au long de la conférence, les pays invités et Statistique Canada ont partagé leurs perceptions et leurs expériences du recensement. Ainsi des questions générales comme celles concernant la confidentialité des données, ou encore des questions plus spécifiques portant sur le développement de l'automatisation, en plus de bien d'autres, ont été soulevées par les conférenciers et les participants, et ont été débattues entre eux. Cette rencontre a donc permis des échanges fructueux entre les pays concernés par les recensements des années 1990.

Ce manuel est divisé principalement en sept séances présentées suivant l'ordre du programme de la conférence. Chaque séance comprend les allocutions qui ont été prononcées ainsi que les résumés des périodes de questions et des discussions qui ont suivi les allocutions. Les périodes de questions et les discussions ont été résumées à partir des bandes sonores enregistrées lors de la conférence. De plus, certaines allocutions ont également été tirées des bandes sonores. Dans ces cas, les textes ont été imprimés en caractère italique afin de ne pas être confondus avec les textes originaux fournis par les conférenciers.

Finalement, deux annexes comprenant le programme de la conférence et la liste des participants ont été ajoutées afin de compléter la documentation sur la conférence.

ALLOCUTION DE BIENVENUE

Ivan P. Fellegi

Statisticien en chef du Canada
Statistique Canada

Observations générales

J'ai le plaisir ce matin de vous souhaiter la bienvenue et d'inaugurer officiellement la présente conférence sur la planification du recensement du Canada de 1991.

La conférence marque le point de départ de nos travaux de planification du recensement de 1991. Nous espérons profiter de l'occasion pour donner un élan à cette entreprise très importante. Les discussions et les échanges d'idées qui auront lieu cette semaine serviront de catalyseur à la plus grande partie des travaux que nous entreprendrons au cours des cinq prochaines années.

Le grand nombre de participants de Statistique Canada et la présence d'un certain nombre de représentants d'organismes statistiques d'autres pays, dont les États-Unis, des pays d'Europe et l'Australie, traduisent l'intérêt que suscite l'objet de la conférence et l'importance qu'elle revêt. Leur contribution et leur expérience seront des plus utiles pour l'établissement des plans que nous élaborerons en vue du recensement de 1991. Il est encourageant de constater le vif intérêt que les participants manifestent, et c'est là un signe très positif.

En un sens, il s'agit en réalité de deux conférences en une. Le Canada compte parmi les rares pays du monde qui procèdent en même temps au recensement de l'agriculture et au recensement de la population. Bon nombre des questions débattues à la conférence s'appliqueront aux deux recensements, mais une séance spéciale sera consacrée au recensement de l'agriculture.

L'importance de la conférence

Le recensement de la population, qui constitue notre programme le plus important et le plus visible, revêt une importance primordiale pour l'appareil statistique du Canada. Le recensement:

- est la seule source de données directement comparables sur les petites régions;
- est la seule source de données sur les groupes de population disséminés;
- fournit les données de référence de bon nombre de nos programmes intercensitaires comme les estimations démographiques, les taux mensuels d'emploi et de chômage et certains éléments du système de comptabilité nationale;
- constitue la principale source de données socio-économiques recoupées (il permet, par exemple, d'établir des rapports concernant la région, l'emploi, l'instruction, l'âge et les liens de parenté).

De la même façon, le recensement de l'agriculture permet d'obtenir le seul ensemble de statistiques régionales exhaustives sur l'agriculture. La précision de notre programme actuel de statistiques sur l'agriculture dépend totalement du recensement.

Cependant, le contexte dans lequel nous effectuons le recensement subit de nombreux changements. Il s'exerce de plus en plus de pression pour que nous réduisions les coûts de réalisation du recensement et, en 1986, nous avons assisté à des modifications sensibles du mode de financement du recensement. Une grande partie du personnel affecté à la collecte et au traitement sera recruté dans le cadre d'un programme d'emploi d'étudiants et de jeunes, et nous tenterons de recouvrer une partie beaucoup plus importante des coûts de réalisation du recensement au moyen d'une nouvelle structure de prix de nos produits et services. Je me doute qu'il s'agit là des mêmes contraintes que celles auxquelles font face les pays représentés ici.

En outre, les techniques de traitement et d'extraction des renseignements évoluent rapidement. Ces nouvelles techniques nous permettent de produire nos données plus rapidement et à un coût

moindre, mais il s'agit d'un couteau à double tranchant. La technologie a aussi pour effet d'accroître la capacité et la demande des utilisateurs des données. Comme ils utilisent une technologie de plus en plus sophistiquée, les utilisateurs exigent aussi des données de qualité supérieure et mettent davantage en question nos méthodes et nos approches.

Les questions comme la protection de la vie privée et le couplage des fichiers de données importants ainsi que la nécessité de poser un grand nombre de questions dans le recensement suscitent également de plus en plus de préoccupations chez le public. La situation qui s'est produite en Allemagne de l'Ouest, loin d'être étrangère à notre contexte, est très inquiétante. En 1991, nous devrons justifier davantage la nécessité de tenir le recensement devant l'opinion publique du Canada.

L'année 1991 peut sembler très éloignée, compte tenu du fait que le recensement de 1986 approche à grands pas. Mais moins de 68 mois nous séparent du mardi 4 juin 1991. Il n'est pas trop tôt pour discuter des questions qui se poseront en 1991. Bon nombre d'autres pays ont commencé à planifier leur recensement de 1990 ou de 1991. Si nous désirons apporter des modifications sensibles au recensement de 1991, il est essentiel que nous commencions à planifier maintenant, et ce même avant de pouvoir tirer profit de notre expérience de 1986.

Nous avons eu récemment une illustration très évidente de l'importance que revêt le recensement au Canada: le recensement de 1986 a été annulé par le gouvernement pour être ensuite rétabli à la suite des arguments convaincants qu'ont fait valoir un grand nombre d'utilisateurs des données. Il vaut la peine de tirer certaines leçons de cette suite d'événements dramatiques.

En premier lieu, le recensement de 1986 a sans aucun doute été rétabli parce que les utilisateurs considéraient le contenu du recensement prévu comme très utile et qu'il répondait à leurs besoins. Il s'agit selon nous de l'objectif le plus important.

En deuxième lieu, aucune réaction négative n'est venue contredire les nombreuses réactions positives que nous avons reçues: c.-à-d. que le recensement avait peu d'ennemis. Il en est ainsi parce que nous avons une bonne réputation au sujet d'un certain nombre de questions-clés: nous protégeons le caractère confidentiel des données; nous avons toujours eu de bons programmes de communication avec le public; nous nous préoccupons de la protection de la vie privée et nous avons diminué continuellement les coûts unitaires par ménage (en dollars constants) au cours des quatre derniers recensements.

En troisième lieu, le rôle des porte-parole influents de l'extérieur a été primordial. Par conséquent, il ne faut ménager aucun effort pour rester en communication étroite avec les principaux utilisateurs et groupes d'utilisateurs.

Attentes vis-à-vis de la conférence

Au cours des trois prochains jours et demi, nous entendrons parler et nous discuterons de nombreux aspects nouveaux et différents de la tenue d'un recensement. Nous nous pencherons notamment sur des questions comme le contenu du recensement, le contexte de la collecte des données qui pourrait exister en 1991 et le recours à l'automatisation. Dans le cadre des débats, nous avons l'intention de jeter les bases de l'établissement de notre stratégie pour 1991. En examinant de nouvelles approches tout en mettant à profit notre expérience de 1981 et de 1986, nous espérons jeter des bases solides pour les futurs travaux de planification.

Nous espérons en outre accomplir certains progrès en vue de l'établissement de nos principales hypothèses de planification du recensement de 1991 et de la détermination de nos priorités en matière de recherche et d'essai. En 1986, notre principale hypothèse de planification était de réduire au minimum la modification du recensement. C'est pourquoi nous avons pu maintenir les coûts unitaires à un niveau très faible en réduisant au minimum la recherche et le développement. Les résultats du recensement de 1986 constitueront l'épreuve décisive de cette hypothèse, mais nous devons commencer maintenant à discuter ce qui, d'après nous, est la façon d'envisager le recensement de 1991 qui convient le mieux.

Enfin, nous espérons sensibiliser le public au recensement de 1991 et encourager une mise en commun des idées au sujet des questions prioritaires à envisager pour réussir à planifier le recensement de 1991. L'expérience de 1981 nous a montré une fois de plus à quel point le recensement est une entreprise qui intéresse tout le Bureau. En 1991, le succès du recensement dépendra sans aucun doute de l'engagement ferme que prendront tous les secteurs de Statistique Canada.

Invités de l'étranger

Nous sommes chanceux de pouvoir compter sur l'aide de plusieurs représentants d'organismes statistiques des États-Unis, du Royaume-Uni, de l'Australie et de la Suède. Trois de ces pays ont assisté à notre conférence sur la diffusion des données du recensement tenue il y a un an, et nous mettons à profit une grande partie de ce que nous avons appris au moment de la planification du programme des produits du recensement de 1986. Par ailleurs, plusieurs pays nous ont dit qu'ils avaient tiré des enseignements de cette conférence.

À l'occasion de la présente conférence sur le recensement de 1991, nous espérons encore une fois tirer parti de l'échange des connaissances et des expériences des représentants d'autres pays. Nous partageons avec eux les défis qui consistent à prévoir les nouveaux besoins en données, à choisir les méthodes de collecte et les techniques de traitement des données et à susciter l'intérêt et l'appui du public à l'égard d'un recensement national. Nous espérons qu'ils tireront grand profit des débats de cette semaine tout comme nous, j'en suis certain, et je leur souhaite la bienvenue.

Conclusion

Pour terminer, je tiens à insister sur le fait que la présente conférence nous offre une possibilité exceptionnelle de réunir les points de vue d'un grand nombre de sommités sur un sujet très important. Les personnes ici présentes constituent une mine de connaissances, d'expérience et d'idées.

C'est en échangeant librement nos idées et nos opinions que nous pourrons le mieux profiter de ces connaissances, et j'encourage tous les participants à participer aux débats. En particulier, je désire encourager les jeunes employés du Bureau et les membres nouvellement arrivés de notre personnel à jouer un rôle actif au cours de cette conférence. Le succès du recensement de 1991 dépend dans une large mesure de vous.

J'aimerais maintenant laisser la parole au président, M. Edward Pryor, directeur général de la Direction de la statistique démographique et du recensement.

ALLOCUTION D'OUVERTURE

Edward T. Pryor

Directeur général de la Direction de la
statistique démographique et du recensement
Statistique Canada

Objet de la conférence

Ce sont nos contacts dans divers pays qui nous ont incités à organiser la conférence. Il y a environ un an, nous avons assisté à une réunion aux États-Unis où nous avons commencé à envisager la planification et la tenue des recensements des années 1990. Il est alors devenu évident qu'il fallait songer au recensement de 1991 même si le recensement de 1986 n'avait pas encore eu lieu.

Les pays qui sont en train de préparer un recensement quinquennal ont souvent tendance à ne pas se préoccuper du recensement décennal pendant la période de pointe du recensement quinquennal. Cela est d'autant plus vrai que notre recensement quinquennal est aussi important que le recensement décennal.

Selon l'ordre chronologique des recensements, il faut d'abord se pencher sur l'élaboration du questionnaire ainsi que sur la collecte, le traitement et la production des données du prochain recensement avant d'aborder le recensement suivant.

Cette façon de procéder pourrait influer sur le recensement décennal de différentes façons et, notamment, du point de vue de l'innovation. De nouvelles questions n'auraient pas eu le temps de surgir, et nous serions obligés de répéter le même contenu que celui du recensement quinquennal. Je suppose que, pour le recensement de 1991, nous ne pourrons pas tolérer ce genre d'immobilisme.

Pour le recensement de 1991, nous devons nous attendre à des modifications et à des changements dans des domaines comme les techniques de collecte, le contenu du recensement et la diffusion des données. De plus, certaines modifications et certains changements pourraient être appliqués d'abord pour le recensement de 1986 afin de permettre de les mettre à l'essai avant leur mise en oeuvre complète.

Échanges internationaux

L'automne dernier, M. Alex Martin, chef de projet du recensement de 1986, a organisé une conférence réunissant plusieurs pays sur la diffusion des données du recensement. Cette conférence était intéressante à plus d'un titre, car différents pays y ont échangé des idées sur le genre de diffusion, et les pays intéressés sont devenus plus conscients du nombre de questions que nous avons en commun. Nous nous sommes rendus compte que nous avons de plus en plus de problèmes semblables à ceux d'autres pays et qu'il serait avantageux pour tous les pays d'en discuter.

Par le passé, les pays qui procèdent à un recensement ont eu tendance à s'isoler et à croire que leurs méthodes de recensement étaient les meilleures. Cependant, en dépit des différences culturelles et des diverses traditions en matière de statistiques, nous avons découvert qu'il serait avantageux de discuter de certaines questions comme la façon de maintenir la comparabilité des données et les changements apportés au contenu avec le temps, la façon de faire le meilleur usage possible de la technologie, le mode de présentation des données, la manière de répondre aux besoins des utilisateurs, etc. De ces découvertes, il en est ressorti un nouveau phénomène que nous pourrions appeler un "esprit de coalition".

Il existe une autre raison de procéder à des échanges internationaux sur les données du recensement. En effet, des organismes comme les Nations Unies, l'Institut international de statistique et l'Union internationale pour l'étude scientifique de la population s'intéressent de plus

en plus à ces renseignements. Ces organismes s'intéressent également aux normes internationales en matière de comparabilité des données. En témoignent les discussions qu'ils ont eues récemment au cours de leurs réunions sur des sujets comme les méthodes de recensement ou le coût de la réalisation d'un recensement. L'intérêt qu'ils ont manifesté devrait nous encourager à poursuivre nos échanges sur les questions relatives au recensement.

En conclusion, j'espère que la conférence et les idées échangées seront fructueuses. En ce qui concerne mes objectifs, je crois que la conférence sera couronnée de succès s'il s'en dégage une bonne idée qui soit originale et nouvelle ou si nous pouvons trouver une nouvelle méthode concernant la collecte des données, l'aspect géographique, le contenu, l'agriculture, le traitement ou les produits du recensement.

J'espère également que nos invités et notre propre personnel prendront connaissance de notions et d'approches qu'ils n'auraient pas pu découvrir s'ils n'avaient pas pu participer à la conférence.

SÉANCE: COLLECTE DES DONNÉES DU RECENSEMENT

Président: John Riddle
Opérations régionales
Statistique Canada

Mardi, le 8 octobre 1985

AUTOMATISATION DE LA COLLECTE DES DONNÉES

BRIAN J. WILLIAMS

OPÉRATIONS RÉGIONALES - MANITOBA ET SUD DE LA SASKATCHEWAN
STATISTIQUE CANADA

Introduction

Le défi des recensements des années 1990 consistera à réduire les travaux qui exigent beaucoup de main-d'oeuvre et à répondre aux pressions des utilisateurs qui souhaitent des données plus actuelles. Les progrès technologiques augmenteront notre capacité d'effectuer la collecte des données, de former le personnel, de contrôler et de surveiller les travaux sur le terrain et le dépouillement, et de stimuler les rapports avec les enquêtés. Le présent document expose un certain nombre de domaines où divers progrès technologiques peuvent faciliter les travaux de collecte des données.

En même temps, les progrès technologiques perturberont encore plus le contexte déjà complexe dans lequel nous devons remplir notre mandat. Les questions d'incursion dans la vie privée qui se posent actuellement prendront encore plus d'importance à mesure que les technologies utilisées pour traiter l'information se perfectionneront. L'illustration de la couverture du rapport annuel de 1984-1985 du Commissaire canadien à la protection de la vie privée montre un bandit masqué qui ouvre furtivement les chaînes d'un micro-ordinateur. L'emploi accru de la technologie peut certainement aider à réduire le coût du dénombrement, mais il soulèvera de nouveaux problèmes et de nouvelles questions du point de vue des recensés.

Collecte des données

La collecte sur le terrain est, de toutes les opérations du recensement, celle qui demande le plus de main-d'oeuvre. Quelque 38,000 recenseurs y travailleront pour le recensement du Canada de 1986. Toute nouvelle technologie qui peut être appliquée à cette phase devrait permettre de réaliser des économies substantielles.

Un domaine auquel on devrait s'attarder est l'utilisation de procédures de contrôle et de suivi téléphonique centralisés dans les centres urbains. Selon un rapport interne de la Division des opérations des enquêtes, daté de mai 1985, le contrôle et le suivi téléphonique centralisés, au lieu de permettre de réaliser des économies, occasionneraient une dépense additionnelle de $100,000[1].

Cette étude supposait que le contrôle et le suivi effectués manuellement par le recenseur seraient remplacés par un contrôle et un suivi manuels centralisés. Cependant, si les opérations de contrôle pouvaient être automatisées au moyen de lecteurs optiques, le contrôle et le suivi téléphonique centralisés deviendraient réalisables. Un questionnaire conçu pour faciliter la reconnaissance optique des caractères avec un système qui permet l'interview téléphonique assistée par ordinateur (ITAO) pour traiter les documents rejetés au contrôle présenterait de nombreux avantages: dépouillement accéléré, relevé automatisé des questionnaires renvoyés, suivi contrôlé. Il convient de souligner que toute modification de cette importance aux procédures de collecte doit faire l'objet d'une étude approfondie et d'essais sur le terrain. Il faut notamment en étudier l'incidence sur la qualité des données. Si le premier contact avec les ménages non répondants était aussi effectué par l'intermédiaire d'un service de suivi centralisé, il est possible que les économies additionnelles ainsi réalisées soient importantes.

Le suivi téléphonique auprès des non-répondants au moyen du système ITAO est aussi réalisable, comme en témoignent en particulier de récentes études portant sur l'enquête sur la population active. L'EPA est l'enquête permanente la plus importante de Statistique Canada puisqu'elle vise environ 55,000 logements chaque mois. Le Bureau utilise depuis peu l'interview téléphonique pour mener cette enquête dans les régions rurales où l'on avait auparavant recours à des interviews sur place. Suite à cette modification dans les procédures, un certain nombre d'études ont été effectuées.

Les résultats préliminaires des deux premiers mois de l'étude portant sur les contacts téléphoniques pour l'enquête sur la population active donnent un taux de dépistage de 70%. Dans cette étude, les logements nouvellement sélectionnés pour l'enquête sont appariés, à partir de l'adresse, à un fichier d'abonnés courants des sociétés de téléphone pour obtenir le nom et le numéro de

[1] Statistique Canada (1983a).

téléphone de l'occupant. Le ménage est alors contacté par téléphone pour répondre à l'enquête sur la population active. Il est évident que cette technique ne peut s'appliquer que dans les grands centres urbains où les dossiers de livraison des questionnaires donnent le numéro de voirie. Néanmoins, dans les plus grands centres régionaux seulement (c.-à-d. Montréal, Toronto, Winnipeg, Edmonton et Vancouver), un suivi téléphonique pourrait être effectué auprès de 325,000 ménages si l'on suppose un taux de réponse par la poste de 85% et un taux de dépistage de 70%. Encore une fois, il est essentiel d'effectuer des essais approfondis sur le terrain. Les conséquences de tout changement de ce genre sont considérables et leur effet pourrait s'étendre à tout le système.

Un autre domaine qui pourrait faire l'objet d'études pour les opérations de collecte en 1991 est l'utilisation d'ordinateurs portatifs dans les régions qui ont été traditionnellement difficiles à dénombrer, par exemple, pour le suivi dans les noyaux urbains. Cette façon de faire permettrait d'accélérer le dépouillement et, grâce à un contrôle instantané, de réduire les erreurs et les risques d'avoir à effectuer une nouvelle visite. Cela permettrait en outre un contrôle plus serré du processus de suivi.

Les utilisations potentielles de ce type particulier d'équipement (ordinateurs portatifs) font ressortir un certain nombre de points importants. Comme organisme, nous avons toujours compté sur la "trace écrite" du document ou du questionnaire. La décision d'employer des ordinateurs portatifs (ou encore un contrôle automatisé avec suivi ITAO) constitue un changement d'approche fondamental. Quelles sont les implications de la perte de la "trace écrite"? Par ailleurs, les ordinateurs portatifs actuellement utilisés causent certaines inquiétudes du point de vue technique. Ceux qui sont vraiment des ordinateurs portatifs ont une capacité de mémoire et d'affichage limitée. Ceux qui ne présentent pas ces contraintes sont portatifs mais ne sont pas faciles à utiliser sur le terrain et leur coût est nettement plus élevé. Il ne fait aucun doute que ces difficultés techniques seront surmontées dans l'avenir. Mais, compte tenu du délai nécessaire pour concevoir les systèmes et procédures, en faire l'essai et les évaluer, les décisions devraient être prises tôt dans le cycle du recensement. Le recensement, de par sa nature même, se prête mal à la recherche et au développement. Il faut privilégier les moyens techniques éprouvés. L'emploi de la technologie est très attrayant et on risque parfois de le considérer comme une panacée. Nos attentes peuvent dépasser les moyens techniques disponibles actuellement.

La technologie présente aussi des avantages pour certaines opérations auxiliaires reliées à la collecte. Dans l'Ouest du Canada, le taux élevé d'exploitants agricoles non résidents entraîne la création de quelque 50,000 formules 6D, Formule de renvoi des terres agricoles (FRTA). Le recenseur doit identifier chaque exploitation agricole dans son secteur de dénombrement (SD) et identifier l'exploitant. Quand le recenseur détermine que l'exploitant habite à l'extérieur du SD ou lorsqu'il ne peut déterminer qui est l'exploitant, il doit établir une formule 6D. Au moyen de ces formules 6D, on retrace l'exploitant et on lui demande de remplir un questionnaire agricole. Or, il arrive fréquemment que les détails qui permettent de retrouver l'exploitant sont incomplets et que le dépistage ne mène nulle part. Il serait très utile de pouvoir accéder aux dossiers automatisés provinciaux sur la taxe foncière, au moyen de la description légale de la terre, afin d'en identifier le propriétaire et, par la suite, l'exploitant.

À mesure que les institutions publiques informatisent leurs dossiers, l'avantage potentiel qui en découle du point de vue des opérations de collecte est immense. L'accès aux banques de données d'autres organismes pour la contre-vérification des dossiers, par exemple, réduirait sans aucun doute les coûts et permettrait d'assurer une meilleure actualité des données. Cette technique, bien qu'elle soit efficace et efficiente, présente un problème juridique. Le Commissaire à la protection de la vie privée, dans son rapport annuel de 1984-1985, décrit l'appariement ou le couplage par ordinateur comme une "menace insidieuse d'envergure pour le mode de pensée et de fonctionnement de notre société"[2].

Finalement, deux des options déjà avancées supposent une combinaison de technologies à la fois pour la collecte et le dépouillement. Nous avons traditionnellement adopté la même approche pour la collecte et le dépouillement. Nous pourrions peut-être employer différentes technologies à l'avenir. Par exemple, il serait possible d'utiliser des lecteurs optiques avec la formule 2A (questionnaire abrégé) puisque le questionnaire n'est pas très long. Autrement dit, la formule 2A pourrait être traitée avec les lecteurs optiques tandis que la formule 2B

[2] Commissaire à la protection de la vie privée (1985).

(questionnaire complet) serait traitée de la façon conventionnelle, c'est-à-dire par l'introduction au clavier.

L'utilisation accrue de la technologie pose un problème d'ordre organisationnel du fait, notamment, que la distinction entre la collecte et le dépouillement est estompée. Actuellement, la collecte et le dépouillement constituent deux phases distinctes. L'utilisation d'ordinateurs portatifs et du contrôle centralisé automatisé entraînera la fusion de la collecte et du dépouillement en une seule opération. La collecte devra être redéfinie comme étant la création d'un fichier de données "sans erreur".

Formation

En 1986, le nombre de recenseurs employés dans les secteurs d'autodénombrement sera de 32,956. La formation de ces recenseurs comprendra huit heures d'autoformation et 6.5 heures de formation en classe. Le coût de cette formation s'élèvera à 2.47 millions de dollars. Les frais de déplacement au titre de la formation feront monter le coût total de la formation à 3.1 millions de dollars. En d'autres termes, chaque minute du programme de formation prévu pour les recenseurs en 1986 coûtera $2,800.

Actuellement, la formation du personnel chargé de la collecte s'effectue presque exclusivement à l'aide de documents imprimés. Il existe plusieurs domaines où l'utilisation d'un matériel relativement standard permettrait de réduire les coûts. Pour le recensement de 1980 aux États-Unis, une série de cassettes audio ont été utilisées pour présenter les processus du recensement au personnel nouvellement embauché dans les bureaux régionaux. Un programme analogue pourrait être utilisé pour l'orientation des chefs de zone de recensement (CZR) et pour la présentation des activités préliminaires. Compte tenu du fait que les CZR sont très dispersés, toute réduction dans la formation au bureau régional permettra de réaliser des économies au chapitre des frais de déplacement. Nous utilisons beaucoup les cassettes audio pour la formation des employés de niveau subalterne dans le cas de nos enquêtes permanentes (c.-à-d. l'enquête sur la population active, l'enquête nationale sur les fermes). Il est généralement reconnu que l'utilisation de cassettes accroît l'efficacité de ces programmes. Il serait possible d'utiliser des cassettes dans les programmes du recensement pour mettre en relief les points essentiels. Ou encore, on pourrait peut-être enregistrer sur cassette de brefs cours de recyclage à l'intention des recenseurs.

Une cassette de ce type insérée dans un répondeur téléphonique permettrait aux recenseurs d'accéder à ce programme de recyclage quand le commissaire au recensement est absent, en composant un numéro de téléphone qui leur aurait été donné à l'occasion de leur cours de formation.

Le secteur le plus dynamique dans le domaine du divertissement à domicile au cours des dernières années a peut-être été celui des enregistreurs de vidéocassettes. Leur nombre dans les foyers canadiens a doublé entre mai 1983 et mars 1984[3] Aujourd'hui, on peut facilement les louer au dépanneur du coin de la rue. Les enregistreurs de vidéocassettes ont été utilisés sur une base ad hoc en 1981 afin de former les chefs de zone de recensement (CZR) et les chefs de district de recensement (CDR) à l'embauchage des commissaires au recensement (CR) ainsi qu'à leurs fonctions en rapport avec les relations publiques. C'est la question de la disponibilité des enregistreurs de vidéocassettes qui a empêché d'en faire un usage général. De plus, il faut disposer de beaucoup de temps pour préparer une bande vidéo si l'on veut que le travail soit bien fait. La bande vidéo est un support inflexible; une modification mineure dans les procédures peut nous obliger à reprendre beaucoup de travail. Néanmoins, certains aspects du programme du recensement pourraient être enregistrés sur bande vidéo. Les enregistreurs de vidéocassettes pourraient être utilisés pour montrer au personnel comment traiter les cas de refus ou comment s'entretenir avec les répondants peu coopératifs; l'enregistreur de vidéocassettes pourrait aussi servir à exposer les responsabilités relatives à la cartographie et d'autres procédures qui sont enseignées moins aisément au moyen d'exercices présentés sur des documents imprimés. Un des problèmes qui peut se poser par suite du pourcentage élevé d'étudiants qui seront embauchés en 1986 est un taux de roulement des recenseurs beaucoup plus élevé que la normale. Pour éviter d'occuper des employés essentiels dans un cycle sans fin de cours de formation, il serait avantageux d'enregistrer sur vidéocassette la formation portant sur la livraison et la reprise. La bande vidéo pourrait être utilisée avec un guide d'autoformation pour former les recenseurs remplaçants.

Avec la collaboration des stations de cablodiffusion communautaire ou des stations éducatives, nous pourrions utiliser les vidéocassettes à une plus grande échelle. Quatre-vingt-dix-huit pour cent des foyers canadiens possèdent au moins un téléviseur noir et blanc[4]. Le programme d'autoformation actuel pourrait être complété par

[3] Statistique Canada (1983b et 1984).
[4] Statistique Canada (1984).

des enregistrements magnétoscopiques reprenant les points principaux, qui seraient diffusés sur les canaux communautaires ou éducatifs en dehors des heures de pointe.

Un autre domaine à étudier est celui de l'utilisation de micro-ordinateurs et (ou) d'ordinateurs portatifs pour aider à la formation du personnel de soutien. On trouve actuellement sur le marché un ordinateur portatif relativement peu dispendieux ($250) conçu pour être utilisé avec un guide d'autoformation. Les modules de mémoire communiquent avec un micro-ordinateur pour produire des données sur l'efficacité des divers aspects de la trousse de formation. Il vaudrait la peine d'utiliser cette application pour vérifier l'efficacité de la formation des recenseurs et (ou) des CR. Cela permettrait de déterminer les sujets qui sont le plus souvent mal compris et d'apporter les améliorations requises pour le recensement suivant. Il est difficile de mesurer l'efficacité de la formation, c'est donc un sujet qui est fréquemment négligé; l'emploi de cette technologie permettrait de résoudre ce problème.

La formation des recenseurs constitue une phase cruciale de l'opération de recensement. Le succès du programme de formation présage, dans une large mesure, du succès du recensement. La technologie nous offrira l'occasion de réduire les coûts de formation tout en augmentant l'efficacité du programme de formation.

Contrôle et surveillance des opérations sur le terrain

Les progrès technologiques nous aideront beaucoup dans de nombreux domaines reliés au contrôle des opérations sur le terrain. Le maintien de lignes de communication avec un personnel nombreux et très dispersé est une tâche difficile mais essentielle. La moindre modification aux procédures après la formation ou le besoin de renseignements additionnels, qui ne peuvent être retirés du Système d'information de la gestion (SIG) nécessite plus de 41,000 appels téléphoniques. Au début des opérations sur le terrain, il arrive souvent que les CZR et les CDR sont en déplacement pour recruter les CR, pour effectuer des contrôles sur le terrain, etc. Les itinéraires sont souvent modifiés pour accommoder les candidats ou à la suite de problèmes soudains. Il peut être difficile de faire parvenir des messages à un CZR. On effectue présentement l'essai d'un système de messagerie électronique vocale qui résoudrait ce problème. "Hello Central" fonctionne comme Envoy 100 ou les autres systèmes de messagerie électronique analogues. Les messages vocaux sont laissés dans une boîte postale électronique à laquelle on peut accéder au moyen d'un téléphone à clavier. On peut accéder à ces messages à partir d'un téléphone à cadran rotatif au moyen d'un simulateur de tonalité. Les frais mensuels sont trois fois moindres que pour un répondeur automatique.

La nouvelle génération de bélinographe accélérera le traitement des rapports du SIG en plus de réduire les coûts. Pour certains des nouveaux appareils, la durée de transmission peut être aussi brève que 30 secondes comparativement à 4 ou 6 minutes dans le cas d'appareils plus vieux. Les jours où jusqu'à 40 CZR doivent transmettre 7 ou 8 rapports, les frais de communication seront grandement réduits et le cumul des données du SIG pourra s'effectuer plus rapidement. En outre, ce qui est essentiel, nous disposerons de plus de temps pour analyser les données et pour réagir aux problèmes, aux niveaux du CZR et du bureau régional.

Le SIG lui-même doit faire l'objet d'une automatisation accrue. Des micro-ordinateurs seront utilisés dans les bureaux régionaux en 1986. À l'avenir, il serait possible d'utiliser des micro-ordinateurs au niveau du chef de zone de recensement (CZR) afin d'accélérer encore la transmission des données du SIG. L'utilisation de micro-ordinateurs à ce niveau permettrait aussi l'élimination des incohérences dans les données du SIG. En outre, le fait d'installer des micro-ordinateurs à ce niveau pourrait aider le CZR à surveiller certaines activités telles que l'embauchage et la formation des recenseurs, le cheminement des demandes de rémunération et des autres documents financiers ainsi que le coût des derniers rappels.

Un goulot d'étranglement notoire qui devrait faire l'objet de plus d'automatisation est le système de rémunération des recenseurs. Au Canada, la production et l'émission des chèques de paye relèvent d'un organisme central, Approvisionnements et Services Canada (ASC). Statistique Canada produit les documents de base pour la paye et les transmet à Approvisionnements et Services qui se charge de l'introduction des données et de l'émission des chèques. Actuellement, le processus est automatisé au niveau d'Approvisionnements et Services Canada, mais non au niveau des BR. Dans les bureaux régionaux, une procédure manuelle compliquée est utilisée pour produire les documents de base destinés à ASC. Ceci a toujours causé des goulots d'étranglement et entraîné de nombreuses plaintes. Le fait de fusionner les éléments du Fichier géographique principal du recensement (FGPR) avec un fichier des recenseurs et les relevés appropriés des documents de travail pour produire un fichier de rémunération pour

introduction dans le système d'ASC permettrait à la fois d'effectuer des économies et de réduire le nombre de plaintes. Nous le savons tous, une seule demande de renseignements présentée par le Ministre peut canaliser beaucoup de temps et d'énergie à différents niveaux de l'organisme.

Relations avec les enquêtés

Dans le domaine de la téléphonie, les nouvelles technologies nous aideront à mieux utiliser le personnel du Service auxiliaire téléphonique (SAT). Exception faite de quelques différences mineures entre les régions, les spécifications actuelles pour le SAT prévoient la disponibilité d'une série de postes téléphoniques à ligne unique reliés à un système de renvoi automatique des appels. La répartition des appels nécessite l'addition d'un matériel dispendieux. Il en va de même dans le cadre de la mise en séquence et de la superposition. Comme le SAT ne fonctionne que pendant un laps de temps très bref, approximativement 10 jours, il n'a pas été pratique d'obtenir ce matériel additionnel. Le distributeur d'appels automatique (DAA) est une caractéristique standard des systèmes centrex de la cinquième génération, qui seront installés dans la majorité des centres au cours des prochaines années. Le DAA combiné à la "superposition" et à la mise en séquence des appels en attente permettra une utilisation maximale du personnel du SAT. Des systèmes de surveillance en direct seront utilisés pour obtenir des renseignements plus détaillés que ce n'est le cas actuellement afin que l'horaire du personnel puisse être modifié pour mieux servir le public.

Un autre domaine qui pourrait faire l'objet d'étude est la possibilité d'établir un système d'appel pour transmettre les données du recensement. Les compagnies d'utilité publique utilisent ce système pour enregistrer les renseignements nécessaires pour la facturation quand l'employé chargé de relever le compteur a trouvé le domicile inoccupé. Cette méthode entraînerait évidemment un certain nombre de problèmes, mais une forme de services d'appels qui permettrait au moins d'enregistrer les rendez-vous ou les heures auxquelles le recenseur peut se présenter chez le recensé permettrait de réaliser des économies dans le cadre du suivi sur place.

Résumé

Des pressions accrues pour réduire les coûts et augmenter l'actualité des données produites forceront Statistique Canada à examiner les possibilités qu'offre la technologie dans le cas des travaux qui exigent beaucoup de main-d'oeuvre. L'application de la technologie aux activités de collecte des données permettrait de réduire les coûts et d'améliorer l'actualité des données, d'améliorer l'efficacité du programme de formation, de coordonner et de diriger plus adéquatement un personnel nombreux sur le terrain. D'un autre côté, la protection de la vie privée restera un sujet de préoccupation dont l'importance grandira vraisemblablement, à mesure que le matériel utilisé sera plus perfectionné. La technologie peut paraître séduisante. Nous devons nous assurer que toute nouvelle technologie utilisée a fait ses preuves. La nature "éphémère" du recensement se prête mal à la recherche et au développement de systèmes qui n'ont pas été essayés. Par ailleurs, il faut disposer d'un délai considérable afin d'acquérir et d'élaborer du nouveau matériel et de nouveaux systèmes. Certains essais sur le terrain seront nécessaires pour évaluer l'effet des nouveautés technologiques sur la qualité des données.

Les domaines que l'on doit sérieusement étudier pour déterminer si l'utilisation de technologies en 1991 est justifiée sont ceux de la formation et de l'emploi des ordinateurs portatifs. Les méthodes actuelles de formation prévoient exclusivement des exercices traditionnels "avec un crayon et du papier". Ceci est non seulement coûteux mais moins efficient qu'une trousse de formation qui incorporerait un nouveau support. Toute réduction d'une heure dans le programme de formation permettrait de réaliser des économies de $168,000. L'élimination d'un cours de formation permettrait de réaliser des économies de $846,000 aux chapitres de la rémunération et des dépenses. L'utilisation d'ordinateurs portatifs permettrait d'accélérer à la fois la collecte et le dépouillement et augmenterait le contrôle exercé sur l'opération.

Les décisions quant à l'obtention d'un nouveau matériel et à l'adoption de nouveaux processus pour 1991 devront être prises prochainement afin de permettre que leur effet soit étudié et évalué de façon détaillée.

BIBLIOGRAPHIE

Barabba, Vincent P., Mason, Richard O. et Mitroff, Ian I., 1983. "Federal Statistics in a Complex Environment. The Case of the 1980 Census", **The American Statistician**, volume 37, nº 3, p. 203-211.

Bounpane, Peter A., 1983. "The Census Bureau Looks to 1990", **American Demographics**, volume 5, nº 10, p. 28-32, 44-48.

General Accounting Office (États-Unis), 1982. **A $4 Billion Census in 1990? Timely Decisions on Alternatives to 1980 Procedures Can Save Millions**, Washington, D.C.

General Accounting Office (États-Unis), 1983. **The Census Bureau Needs to Plan Now for a More Automated 1990 Census**, Washington, D.C.

Commissaire à la protection de la vie privée, 1985. **Rapport annuel du Commissaire à la protection de la vie privée, 1984-1985**, Ottawa.

Robey, Bryant, 1983. "Achtung! Here Comes the Census", **American Demographics**, volume 5, nº 10, p. 2-4.

Robey, Bryant, 1984. "The 1990 Census: A View from 1984", **American Demographics**, volume 6, nº 7, p. 24-29, 46.

Statistique Canada (non publié), mai 1983. **Centralized Edit and Telephone Follow-up Study**, Ottawa.

Statistique Canada, mai 1983. **Équipement ménager**, nº 64-202 au catalogue, Ottawa.

Statistique Canada, mars 1984. **Équipement ménager**, nº 64-202 au catalogue, Ottawa.

SUIVI TÉLÉPHONIQUE - PASSÉ, PRÉSENT ET AVENIR

JOHN A. KAZMAIER JR.

FIELD DIVISION
U.S. BUREAU OF THE CENSUS

Introduction

Le plan de collecte des données du recensement décennal de 1980 et du recensement d'essai de 1985 comprenait un contrôle des questionnaires retournés par la poste en vue d'évaluer dans quelle mesure les réponses étaient précises et complètes. Les questionnaires jugés incomplets ont fait l'objet d'un suivi téléphonique ou d'une visite sur place afin d'obtenir les renseignements manquants.

Notre but, lors des deux recensements, était de réduire le coût des opérations sur le terrain sans nuire à la qualité des données. Le suivi téléphonique est un moyen pour réduire le suivi sur place et, par là même, les coûts. Nous continuerons à effectuer des suivis téléphoniques au cours des prochains recensements.

Expérience tirée du recensement de 1980

Méthodologie du recensement pour le système de contrôle et de suivi

- Les questionnaires retournés par la poste et les questionnaires remplis par les recenseurs ont fait l'objet d'un contrôle manuel.

- Les questionnaires jugés incomplets lors du contrôle ont été envoyés soit pour suivi téléphonique, soit pour suivi sur place[1]. Ces questionnaires comprenaient les questionnaires complets pour lesquels 20 questions ou plus étaient sans réponse et les questionnaires abrégés pour lesquels quatre questions ou plus étaient sans réponse. Le suivi téléphonique s'est révélé plus efficace pour ce groupe de questionnaires sans réponse.

- Les commis chargés du suivi téléphonique ont été entraînés à entrer en communication avec les ménages et à poser à nouveau les questions jugées incomplètes au contrôle. Lorsque le recensé n'avait pas inscrit de numéro de téléphone sur son questionnaire, le commis chargé du suivi téléphonique devait utiliser les annuaires téléphoniques et les annuaires rue-adresse pour obtenir le numéro de téléphone. En dernier recours, il pouvait appeler l'assistance-annuaire pour obtenir le numéro de téléphone. Si le commis ne pouvait contacter le ménage après cinq tentatives ou s'il ne pouvait communiquer avec un répondant acceptable, les questionnaires étaient mis de côté pour une interview sur place.

- Des commis au contrôle qualitatif se sont assurés que toutes questions relevées au cours du contrôle avaient été remplies de façon satisfaisante par les commis chargés du suivi téléphonique ou que le questionnaire avait été marqué pour suivi sur place. Si un questionnaire marqué pour suivi sur place contenait moins de quatre questions sans réponse (moins de 20 dans le cas des questionnaires complets), le cas était considéré comme réglé.

Dotation et horaire

Approximativement 4,000 lignes téléphoniques étaient prévues au budget pour le suivi téléphonique à l'échelle nationale; environ 950 de ces lignes avaient été utilisées auparavant pour aider les recensés à remplir leur questionnaire par téléphone. En 1980, le suivi téléphonique a été effectué à partir de 87 bureaux centralisés, dans tout le pays.

En 1980, le suivi téléphonique centralisé a commencé approximativement une semaine après le début de l'opération de contrôle. Il s'est terminé environ une semaine après la fin du contrôle.

Le service téléphonique fonctionnait six jours par semaine, du lundi au samedi.

[1] Dans les bureaux de district centralisés, le suivi téléphonique des questionnaires incomplets s'est effectué à partir du bureau central. Dans les bureaux de district décentralisés, les questionnaires incomplets ont été remis à des recenseurs qui ont communiqué avec les ménages au moyen de leur téléphone privé ou en effectuant un suivi sur place.

Dans les bureaux centralisés, l'horaire de travail des employés était le suivant:

Poste de jour

1. Surveillant des opérations de bureau

   Du lundi au vendredi -- de 8 h 30 à 17 h 00
   Samedi -- de 8h 30 à 13 h 15

   Le surveillant des opérations du bureau ou son adjoint travaillait les samedis après-midi, au choix de la personne.

2. Commis de bureau principaux

   Du lundi au samedi -- de 8 h 15 à 15 h 15 (avec une pause d'une demi-heure pour le repas)

3. Commis au contrôle qualitatif

   Du lundi au samedi -- de 8 h 15 à 15 h 15 (avec une pause d'une demi-heure pour le repas)

4. Commis préposés au suivi téléphonique

   Du lundi au vendredi -- de 8 h 30 à 15 h 00 (avec une pause d'une demi-heure pour le repas)

Poste de nuit

1. Adjoint au surveillant des opérations de bureau

   Du lundi au vendredi -- de 13 h 15 à 21 h 45
   Samedi -- de 13 h 15 ou de 17 h 00 à 21 h 45 (selon l'horaire du surveillant des opérations de bureau)

2. Commis de bureau principaux

   Du lundi au vendredi -- de 14 h 45 à 21 h 45 (avec une pause d'une demi-heure pour le repas)

3. Commis au contrôle qualitatif

   Du lundi au samedi -- de 14 h 45 à 21 h 45 (avec une pause d'une demi-heure pour le repas)

4. Commis chargés du suivi téléphonique

   Du lundi au samedi -- de 15 h 00 à 21 h 30 (avec une pause d'une demi-heure pour le repas)

Résultats de l'évaluation

Une évaluation a été faite afin de déterminer l'efficacité du système de contrôle et de suivi utilisé en 1980. En ce qui a trait au suivi téléphonique des questionnaires retournés par la poste, les conclusions sont les suivantes:

1. Dans les bureaux centralisés, le suivi téléphonique n'a résolu que 26% des questions relatives à la population et au logement pour lesquelles les commis au contrôle avaient indiqué qu'un suivi devait être effectué[2].

2. Le suivi téléphonique a donné des résultats de beaucoup supérieurs dans le cas des questions posées à l'ensemble de la population que dans celui des questions posées à un échantillon de ménages. Le taux de réussite a été de 37.2% et de 27.2%, respectivement.

3. Dans les bureaux centralisés le suivi téléphonique n'a pas donné les résultats escomptés, bien que ceux-ci aient été supérieurs aux résultats des suivis sur place. Voici un tableau des résultats:

QUESTIONS SANS RÉPONSE

Au début	Après le contrôle par les commis
23.0%	16.9%
Après le suivi téléphonique	**Après l'interview sur place**
12.1%	10.0%

   La conclusion tirée de cette analyse est que le suivi sur place n'a pas réduit les cas de non-réponse de façon substantielle après le suivi téléphonique.

   Bien que le suivi téléphonique ait permis de réduire d'environ 25% le nombre de questionnaires sans réponse, cela ne veut pas dire que, dans le cas des questions pour lesquelles il y avait le plus de cas de non-réponse au début, la réduction du taux de non-réponse a été aussi élevée.

4. Le coût d'un suivi téléphonique est d'environ $1 par questionnaire comparativement à $4 pour le suivi sur place.

[2] La base utilisée pour déterminer ce taux comprend les questionnaires incomplets jugés acceptables d'après les règles de tolérance relatives au contrôle.

Expérience tirée des recensements d'essai de 1985

Méthodologie du recensement pour le système de contrôle et de suivi

- Les questionnaires retournés par la poste et les questionnaires remplis par les recenseurs à partir des sites où la collecte des données est effectuée (Jersey City, New Jersey et Tampa, Floride) ont été dépouillés à notre centre de dépouillement permanent à Jeffersonville, Indiana.

- Un contrôle effectué par ordinateur a identifié les questionnaires rejetés au contrôle.

- Une sous-section de revue manuelle a tenté de corriger les erreurs qui avaient entraîné le rejet des questionnaires au contrôle.

- Les questionnaires retournés par la poste pour lesquels il a été impossible de corriger les erreurs ont été envoyés à une sous-section de téléphonistes pour le suivi.

- Les commis chargés du suivi téléphonique ont été entraînés à entrer en communication avec les enquêtés et à poser à nouveau les questions jugées incomplètes au contrôle.

- Si un commis chargé du suivi téléphonique ne pouvait entrer en communication avec un répondant après trois tentatives ou si un répondant demandait un suivi sur place, le questionnaire original retourné par la poste était envoyé au bureau chargé de la collecte des données en vue d'un suivi sur place.

Dotation et horaire

Les opérations de suivi téléphonique ont duré environ deux semaines.

Une seule équipe a été employée. Le poste de travail allait de 11 h 30 à 20 h 00 (du lundi au vendredi).

L'équipe de travail se composait des employés suivants:

1. Surveillant principal

2. 2 ou 3 opérateurs chefs (surveillants adjoints)

3. 55 commis (bien que 60 téléphones aient été installés, tous n'ont pas été utilisés).

Résultats de l'évaluation préliminaire

Aucune évaluation explicite du suivi téléphonique relatif au recensement d'essai de 1985 n'a été effectuée. Cependant, certains résultats préliminaires sont disponibles:

1. Dans approximativement 80% des cas de questionnaires retournés par la poste et rejetés au contrôle, un numéro de téléphone était fourni (soit qu'il ait été inscrit par le recensé sur le questionnaire, soit qu'il ait été obtenu par un commis à Jeffersonville à partir des annuaires téléphoniques ou de l'assistance-annuaire).

2. Les commis chargés du suivi téléphonique ont communiqué avec 78.7% des ménages dont les questionnaires ont été rejetés au contrôle.

   (NOTA: La majorité des commis chargés du suivi téléphonique étaient des nouveaux employés. Cependant, nous avons aussi employé quelques personnes qui avaient l'expérience des interviews téléphoniques.)

Avantages potentiels du suivi téléphonique

1. L'expérience a démontré que, dans le cas de questionnaires rejetés au contrôle, le coût unitaire du suivi est beaucoup moindre pour les suivis téléphoniques que pour les suivis sur place. C'est une des principales raisons pour lesquelles le suivi téléphonique est utilisé le plus possible dans le cas des questionnaires rejetés au contrôle. Dans les bureaux décentralisés, en 1980, les questionnaires rejetés au contrôle ont été envoyés aux recenseurs. Ces derniers devaient utiliser leur téléphone privé afin de régler le cas des questionnaires pour lesquels un numéro de téléphone était fourni, soit qu'il soit indiqué sur le questionnaire, soit qu'il se trouve dans un annuaire téléphonique.

2. L'accent étant mis sur l'utilisation maximale du téléphone, le Census Bureau effectuera aussi l'essai de méthodes pour effectuer un suivi téléphonique dans les cas de non-réponse au cours des recensements d'essai de 1986, en plus d'utiliser le téléphone pour régler les cas des questionnaires rejetés au contrôle.

   Le suivi des cas de non-réponse consiste à communiquer avec les ménages qui n'ont pas renvoyé par la poste leur questionnaire rempli et à les interviewer. Les listes de tâche fournies aux recenseurs contiendront toutes les adresses des ménages pour lesquels il faut remplir un questionnaire. Certaines de ces

adresses inclueront aussi des numéros de téléphone. Les recenseurs seront formés pour utiliser les numéros de téléphone afin d'obtenir des interviews complètes. Si le questionnaire d'un ménage est rempli par téléphone, le recenseur doit indiquer, sur le questionnaire même, qu'il s'agit d'un cas réglé par téléphone.

Certains numéros de téléphone seront supprimés des tâches relatives au suivi des cas de non-réponse. Ces cas seront utilisés comme groupe de contrôle pour effectuer la comparaison des données avec les cas réglés par téléphone. De plus, des mesures de la couverture (par couverture on entend ici le dénombrement des bonnes personnes aux bons endroits, sans répétition) seront effectuées à la suite des interviews sur place aux fins du suivi.

Par ailleurs, des plans sont ébauchés pour évaluer le traitement des cas par téléphone uniquement, c'est-à-dire que certains des cas rejetés au contrôle ne seront jamais retournés sur le terrain pour un suivi sur place mais feront l'objet d'un autre suivi téléphonique. D'autres cas rejetés au contrôle seront réattribués en vue d'un suivi sur place. On procédera alors à une comparaison des questions laissées sans réponse et des coûts des suivis téléphoniques additionnels par rapport au coût des suivis sur place.

3. Le Census Bureau utilise l'interview téléphonique assistée par ordinateur (ITAO) pour certaines de ses enquêtes sporadiques ou permanentes. Une installation d'ITAO a été établie à Hagerstown, Maryland.

   Si le Census Bureau effectue un suivi téléphonique assisté par ordinateur des questionnaires rejetés au contrôle pour le recensement, il se peut qu'il soit possible de réduire encore le nombre de questions laissées sans réponse. Un programme de traitement assisté par ordinateur force l'opérateur de terminal ou le commis affecté au téléphone à introduire de nouvelles données pour les questions laissées en blanc, sans avoir à parcourir un document imprimé pour trouver toutes les questions qui doivent faire l'objet d'un suivi. Un tel système permet aussi de concevoir des questions supplémentaires pour obtenir les renseignements manquants. Ainsi, la conception des procédures opérationnelles pourrait être plus adéquate dans le cadre d'un programme assisté par ordinateur.

   On pourrait également avoir recours à un système de gestion automatisé des appels pour accroître l'efficacité de la phase des rappels. Le système comprendrait une base de données regroupant tous les cas pour lesquels il y a eu rejet au contrôle, et il permettrait de traiter ces derniers à certaines heures de la journée, compte tenu du relevé des contacts antérieurs.

4. L'évaluation du recensement de 1980 a indiqué que le taux de règlement des cas de non-réponse à la suite du suivi téléphonique était supérieur au taux de règlement des cas rejetés au contrôle qui ont fait l'objet d'un suivi sur place. Notre conclusion est donc que le suivi téléphonique constitue une opération valable.

Difficultés relatives au service de soutien et aux activités opérationnelles

Certains des problèmes mentionnés ci-dessous ont été relevés au cours de recensements antérieurs, les autres sont ceux qui pourraient se poser suite à la modification du contexte.

1. De nombreux recensés ne veulent pas inscrire leur numéro de téléphone sur le questionnaire du recensement. De plus, il y a eu une augmentation du nombre de numéros de téléphone confidentiels ou non publiés dans certaines régions.

   Le public semble être plus soupçonneux et un nombre croissant de personnes refusent de fournir au téléphone des renseignements à propos d'elles-mêmes ou de leur domicile. Les sociétés privées utilisent de plus en plus le téléphone pour trouver des clients pour leurs produits ou présenter leur publicité. Nombre de recensés sont devenus hostiles à ce mode de collecte de données à cause des vendeurs qui cherchent à rejoindre une clientèle par téléphone. Dans certains cas, l'employé du Census Bureau qui effectue une interview téléphonique est perçu comme une personne qui fait de la sollicitation par téléphone.

2. Depuis que AT&T a perdu son monopole, nous pourrions faire face à des difficultés considérables pour l'installation de raccordements ou de lignes WATS dans un emplacement centralisé, particulièrement si le suivi téléphonique est très centralisé.

   Avec la déréglementation, le coût des lignes téléphoniques et de l'installation des appareils pourrait être beaucoup plus élevé dans les emplacements centralisés que ce ne fut le cas en 1980.

3. S'il faut effectuer des recherches en rapport avec les numéros de téléphone et si les annuaires téléphoniques et les annuaires rue-adresse sont les seuls documents de référence, il faudra consacrer beaucoup de temps à trouver le bon numéro de téléphone. Il est très difficile de trouver un numéro de téléphone dans le cas de personnes qui habitent dans des immeubles à logements multiples, particulièrement quand rien ne permet d'identifier le logement (par exemple, app. 1 ou app. A). Les commis doivent consacrer beaucoup de temps à apparier les noms et les adresses. Il se peut que le dépistage par ordinateur soit une option viable, mais beaucoup de recherches reste à faire dans le domaine de l'appariement des adresses par ordinateur.

4. L'expérience acquise lors du recensement de 1980 révèle que les opérations ne se sont pas toujours déroulées de façon uniforme pendant le suivi téléphonique. Les commis chargés du suivi téléphonique ont manqué de travail mais les opérations qui précèdent le suivi (contrôle et contrôle qualitatif du contrôle) n'ont parfois pas produit de travail additionnel au bon moment pour occuper les commis. Ceci a eu un effet sur les coûts, du fait que la répartition du travail n'a pas toujours été efficace et que le personnel n'a pas été régulièrement occupé. Cependant, ce problème disparaît quand on a recours à l'automatisation (contrôle informatisé/ITAO).

Objectifs et buts du Census Bureau pour un suivi téléphonique réussi

1. Le Census Bureau s'est engagé à effectuer le recensement rapidement à un coût unitaire relativement faible tout en maintenant une qualité des données élevée. Compte tenu des contraintes budgétaires et des délais fixés, nous devrons utiliser le suivi téléphonique dans tous les cas où cela est possible. Toute opération de suivi téléphonique doit être conçue de façon à réduire lé plus possible le nombre de questions laissées sans réponse dans les questionnaires rejetés au contrôle.

   Une façon d'augmenter les chances de succès serait de concevoir le questionnaire afin qu'il soit plus facile à utiliser. Certains commis chargés du suivi téléphonique ne pouvaient identifier facilement les questions qui devaient faire l'objet du suivi, à cause de la présentation du questionnaire du recensement de 1980.

2. Il est logique de régler les cas rejetés au contrôle à l'endroit où les questionnaires renvoyés par la poste sont reçus. Lorsque le document lui-même est utilisé pour entrer en communication avec le ménage et pour régler le cas des questions rejetées au contrôle, il est alors pratique du point de vue du personnel de soutien et du personnel opérationnel que le suivi téléphonique centralisé soit effectué au moyen des questionnaires renvoyés par la poste. Autrement, il faudrait transmettre les questionnaires rejetés par la poste à plusieurs centaines de bureaux chargés de la collecte des données (et à des centaines d'employés) pour régler les cas rejetés au contrôle. Ces questionnaires devraient ensuite être renvoyés au bureau central pour la mise à jour des données.

   Cependant, il se peut qu'il soit impossible d'installer des centaines de lignes téléphoniques à un seul endroit aux fins du suivi téléphonique, à cause du manque de lignes téléphoniques raccordées à un immeuble particulier. De plus, le nombre de lignes est limité dans certaines zones de rattachement.

3. Il serait possible d'utiliser les différentes applications du téléphone assisté par ordinateur, comme un système de gestion automatisé des appels, pour augmenter l'efficience des opérations.

   Cependant, il peut ne pas être rentable d'installer un réseau de terminaux d'ordinateurs pour chaque commis chargé du suivi téléphonique. En 1980, 7,500 commis ont été employés pour le suivi téléphonique. Comme l'opération n'a duré que de 3 à 4 semaines, il ne serait probablement pas pratique de relier plusieurs centaines de terminaux d'ordinateurs à un (ou plusieurs) mini-ordinateur pour effectuer des interviews téléphoniques assistées par ordinateur.

   Tout système de gestion automatisé des appels devra faire l'objet d'essais pour en vérifier la faisabilité. Si l'ITAO n'est pas utilisée, il pourrait être très difficile d'avoir recours à un système de gestion automatisé des appels. Les données devraient être introduites à partir des relevés des appels, puis quelqu'un devrait affecter manuellement les cas rejetés au contrôle non réglés à diverses équipes de travail selon les résultats du tri effectué par le système de gestion automatisé des appels. Ce dernier système serait probablement géré par un ordinateur individuel.

Si nous utilisons un système automatisé de collecte des données sur le terrain avec des ordinateurs portatifs, un système de gestion automatisé des appels pourrait alors être viable.

***Nota**:*

Recensement d'essai des É.-U. de 1986

*Au cours de son exposé, **M. Kazmaier** a discuté plus amplement du recensement d'essai de 1986 qui sera tenu à Central Los Angeles (Californie) auprès d'environ 240,000 logements, et dans huit comtés de East Central Mississippi auprès d'environ 80,000 logements.*

Les méthodes de contrôle et de suivi employées aux deux endroits seront semblables. Les questionnaires de retour par la poste qui ne peuvent être complétés par suivi téléphonique seront retournés au bureau de la collecte en vue d'un suivi sur place.

Ces recensements d'essai de 1986 visent non seulement à mettre à l'épreuve le recours au suivi téléphonique dans les cas de questionnaires rejetés au contrôle mais aussi pour les cas de non-réponse.

Une liste des numéros de téléphone pour la région d'essai à Los Angeles a été appariée au fichier de contrôle des adresses. Cet appariement est effectué à l'aide du Fichier de base géographique (FBG), fichier de cartes de rues lisibles par machine. Les recenseurs devront utiliser la liste des numéros de téléphone pour effectuer le suivi téléphonique dans les cas de non-réponse.

Au Mississippi, il a été impossible d'apparier les numéros de téléphone avec le fichier de contrôle des adresses, étant donné que le FBG ne couvre pas cette région. Par conséquent, les recenseurs devront d'abord procéder au suivi sur place, puis, si ce dernier se révèle infructueux, effectuer le suivi téléphonique après avoir obtenu le numéro de téléphone approprié par l'entremise des voisins ou d'un annuaire téléphonique.

Enfin, les résultats des recensements d'essai de 1986 permettront de détérminer si le recours au suivi téléphonique dans les cas de questionnaires rejetés au contrôle et dans les cas de non-réponse a un effet sur la qualité des données et la précision de la couverture.

EXTENSION DU RETOUR PAR LA POSTE AUX SECTEURS DE REPRISE

DOUG HICKS

DIVISION DES OPÉRATIONS DES ENQUÊTES
STATISTIQUE CANADA

Introduction

Dans le recensement actuel, 99% des ménages et des exploitations agricoles au Canada sont dénombrés par autodénombrement. Il y a deux formes d'autodénombrement: le retour par la poste[1], utilisé dans les grandes régions urbaines, et la reprise[2], utilisée dans les petites régions urbaines et les régions rurales. Le présent document décrit les implications de l'adoption universelle du retour par la poste en 1991. Nous étudierons en particulier la rentabilité potentielle, la réduction dans les activités qui emploient beaucoup de main-d'oeuvre, les difficultés qui pourraient en résulter sur le plan opérationnel, l'établissement des tâches, l'effet sur le recensement de l'agriculture ainsi que d'autres aspects de cette mesure.

1. Rentabilité et activités qui emploient beaucoup de main-d'oeuvre pour le recensement de la population

Le tableau 1 fait état des économies qui pourraient être réalisées si le retour par la poste était employé dans les secteurs de reprise, sans modification dans la taille des tâches par rapport à 1986.

Tableau 1: Comparaison des économies hypothétiques du retour par la poste par rapport à la reprise, Canada, recensement de 1986

Désignation géographique		Taux pour le retour par la poste	Taux pour la reprise	*Économies par ménage	Nombre estimatif de ménages	Économies
		($)	($)	($)		($)
A	2A	1.30	1.71	.032	960,000	31,000
	2B	1.98	2.36			
B	2A	1.43	1.88	.070	230,000	16,000
	2B	2.12	2.53			
C	2A	1.66	2.19	.144	400,000	58,000
	2B	2.38	2.84			
D	2A	2.00	2.64	.246	850,000	210,000
	2B	2.75	3.28			
E	2A	3.17	4.18	.592	285,000	170,000
	2B	4.04	4.82			
						485,000

* Les économies moyennes par ménage selon la désignation géographique sont déterminées à l'aide de la formule suivante:

$$\frac{((TR2A - TrP2A) \times 4 + (TrP2B)) - ((TP2A \times 4 + TP2B) \times TrP\,90\%)}{5}$$

où
TR: taux de réponse pour une formule ___
TrP: taux pour le retour par la poste pour une formule ___ ou taux de retour par la poste de ___%
TP: tarif postal pour une formule ___

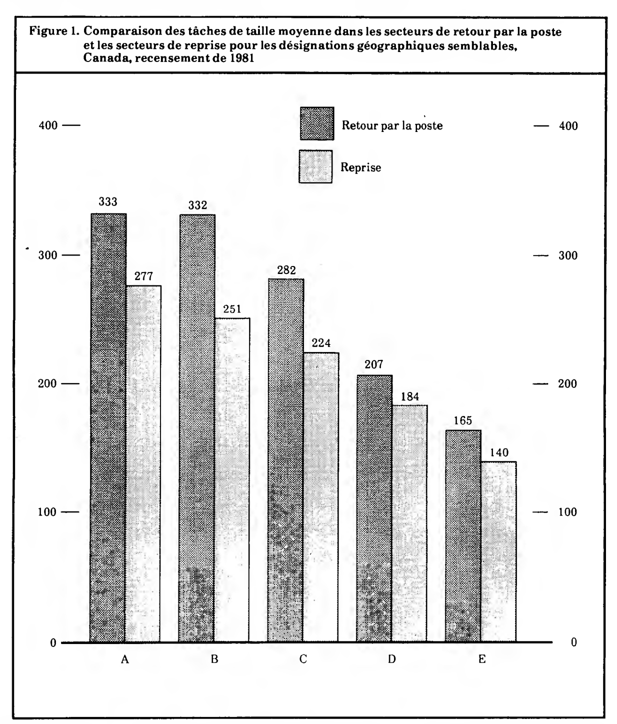
[1] Le retour par la poste est défini comme le renvoi d'un questionnaire, par la poste, au recenseur pour chaque logement privé où un questionnaire a été livré. Le recenseur contrôle les questionnaires renvoyés et effectue un suivi téléphonique et (ou) un suivi sur place pour les cas de questionnaires rejetés au contrôle et pour les cas de non-réponse.

[2] La reprise est définie comme une nouvelle visite effectuée par le recenseur à chaque logement privé pour reprendre et contrôler tous les questionnaires livrés.

Nous avons toute raison de croire, à mesure que nous approchons de 1991, que les économies réalisées en optant pour le retour par la poste au lieu de la reprise seraient même supérieures. On s'attend à ce que les augmentations dans le coût de la main-d'oeuvre et dans les frais de déplacements relatifs à la reprise dépassent toute augmentation dans les tarifs postaux relative au retour par la poste. Le succès de la reprise repose sur la possibilité de rencontrer le recensé pour récupérer le questionnaire. Plus le pourcentage de cas "Aucun contact" est élevé, plus le nombre des visites de rappel augmente, ce qui accroît les coûts liés à la récupération des questionnaires. Ce problème, ainsi que les coûts correspondants, peut être minimisé en optant pour le retour par la poste.

Le graphique qui suit compare des tâches de taille moyenne en 1981, selon la méthode utilisée, pour des désignations géographiques semblables.

Figure 1. Comparaison des tâches de taille moyenne dans les secteurs de retour par la poste et les secteurs de reprise pour les désignations géographiques semblables, Canada, recensement de 1981



Des économies seraient réalisées grâce à une diminution des besoins en main-d'oeuvre si la taille des tâches dans les secteurs de reprise était proche de celle des tâches dans les secteurs de retour par la poste. Une réduction du nombre de recenseurs a de nombreuses répercussions. En effet, l'abolition de 15 à 20 postes de recenseur provoque la suppression d'un poste de commissaire au recensement, d'un poste d'adjoint administratif au commissaire au recensement, et l'équivalent d'une partie d'un poste de technicien du contrôle qualitatif. En 1986, l'abolition d'un poste de recenseur permettrait d'épargner entre $150 et $200, d'un poste de commissaire au recensement, entre $6,500 et $7,000, d'un poste d'adjoint administratif au commissaire au recensement, entre $550 et $600, et d'un poste de technicien du contrôle qualitatif, entre $700 et $800. Si la taille des tâches dans les secteurs de reprise actuels équivalait à peu près à celle des tâches dans les secteurs de retour par la poste, il serait sans doute possible de retrancher 1,500 personnes des effectifs et d'économiser ainsi $750,000.

D'un autre côté, l'emploi universel de la méthode de retour par la poste entraînerait des dépenses supplémentaires, notamment:

- la livraison des questionnaires aux recenseurs effectuée à partir des bureaux des commissaires au recensement;
- des frais interurbains éventuels pour le suivi téléphonique dans certaines régions;
- une augmentation des coûts de production attribuable à la production d'une plus grande quantité d'enveloppes de retour par la poste.

Au total, ces dépenses ne devraient pas dépasser $300,000.

2. Considérations d'ordre opérationnel

2.1 Système téléphonique

Le succès de l'élargissement de la méthode du retour par la poste repose sur la possibilité d'obtenir un taux de réponse élevé et d'effectuer un suivi téléphonique pour les questionnaires rejetés au contrôle et, éventuellement, pour les ménages non répondants. Toutefois une restriction s'applique, dans le cas du suivi téléphonique, en ce qui a trait aux régions rurales où un nombre important d'abonnés partagent les lignes téléphoniques.

D'après les données de 1980-1981, on estime à 50% la proportion de lignes privées, ce qui donne un pourcentage équivalent de ménages utilisant des lignes partagées.

Il semble que la popularité des lignes partagées soit à la baisse depuis 1980 et que cette baisse continuera de s'accentuer d'ici 1991. Néanmoins, on s'attend à ce qu'en 1991 il y en ait encore un nombre important dans les régions rurales ce qui présentera un risque du point de vue de la confidentialité des données si l'on maintient les mêmes procédures de suivi dans les secteurs de retour par la poste. Cet aspect devra être dûment pris en considération pour que l'élargissement de la méthode du retour par la poste puisse être considéré comme une option valable. Il ne semble pas que cela constitue un problème insurmontable étant donné qu'actuellement beaucoup d'enquêtes se font par téléphone et que cette tendance va s'accentuer à l'avenir.

2.2 Système postal rural

Le succès de l'élargissement de la méthode de retour par la poste dépend également des délais de réception et de livraison des questionnaires retournés par la poste. Trois facteurs influent sur le nombre de questionnaires reçus par les recenseurs:

- le degré de coopération du public, c'est-à-dire la bonne volonté avec laquelle celui-ci consent à remplir son questionnaire et à le retourner par la poste;
- la possibilité pour Postes Canada de traiter les questionnaires retournés et de les transmettre aux commissaires au recensement;
- l'efficacité avec laquelle les commissaires au recensement trient les questionnaires reçus et les transmettent aux recenseurs.

2.2.1 Coopération du public

Étant donné les résultats du recensement de 1981 et le taux de questionnaires retournés par la poste enregistré à l'occasion du recensement d'essai de 1977, il est fort probable que l'élargissement de la méthode du retour par la poste sera caractérisé par des taux de réponse tout à fait acceptables.

2.2.2 Systèmes postaux canadiens dans les régions rurales

L'accroissement du nombre de secteurs de retour par la poste complique quelque peu le processus d'acheminement des questionnaires retournés par la poste. Dans la plupart des secteurs de reprise actuels, il est impossible pour les recensés de jeter leur questionnaire dans la boîte aux lettres la plus proche. Le questionnaire doit être récupéré par une compagnie de messagerie locale ou encore le recensé doit se rendre au bureau de poste voisin pour l'expédier. Du bureau de poste local, les questionnaires doivent alors être acheminés au bureau de poste de la région où se trouve le commissaire au recensement. D'après une étude restreinte effectuée par le personnel de Postes Canada dans les districts de commissaire au recensement où des problèmes risquent de se poser, le délai d'acheminement des questionnaires pourrait aller de 2 à 4 jours. Le fait que la réception des questionnaires pourrait être légèrement retardée ne saurait être une raison suffisante pour renoncer à étendre la méthode du retour par la poste. Bien que cela présente quelques inconvénients du point de vue de l'opération sur le terrain, il est possible de s'accommoder de tels délais.

2.2.3 Distribution par les commissaires au recensement des questionnaires retournés par la poste

La distribution opportune des questionnaires aux recenseurs, par les commissaires au recensement entraînera vraisemblablement des dépenses additionnelles. Il conviendrait d'évaluer les calendriers et les plans actuels et de faire des recommandations en vue de minimiser l'effet qui en résultera sur les coûts. On pourrait également envisager la possibilité que les questionnaires soient remis par le personnel du bureau de poste directement aux recenseurs.

2.3 Délai de collecte

Si le projet d'élargissement de la méthode du retour par la poste est maintenu, il faudra étudier en détail l'effet qui en résultera sur le délai de collecte des données. L'essai portant sur l'élargissement de cette méthode, effectué en 1977, a révélé qu'il s'agit d'une alternative très intéressante, mais n'a nullement conduit à la conclusion qu'elle était préférable à la reprise. Les facteurs qui sont apparus comme étant susceptibles d'occasionner des problèmes et de rallonger le processus de collecte des données devront être étudiés afin d'en réduire le plus possible les effets négatifs.

3. Établissement des tâches

Même si elle dépasse le simple cadre de l'élargissement de la méthode du retour par la poste, la question de l'établissement des tâches est abordée ici parce que, pour que des économies puissent être réalisées, une telle mesure devrait être assortie d'une révision des limites des secteurs de reprise.

Actuellement, chaque recenseur est chargé du dénombrement d'une région géographique définie, désignée comme secteur de dénombrement. En raison des caractéristiques géographiques, du coût d'une révision des limites et du désir au cours des derniers recensements de préserver la comparabilité des données chronologiques en assurant la stabilité des limites, il est difficile de bénéficier des économies qu'un accroissement des charges de travail permettrait de réaliser.

Plus la révision des limites se rapprochera de la taille idéale des tâches, plus les économies réalisées en termes de ressources humaines et financières seront grandes. Cela vaut non pas uniquement dans le cas de l'élargissement de la méthode du retour par la poste mais également pour toute autre méthode de collecte.

La solution la plus intéressante consiste à opter pour une certaine flexibilité dans l'établissement des tâches. Il faudrait pouvoir adapter la taille des tâches en fonction des conditions d'efficience particulières à chaque méthode de collecte.

4. Recensement de l'agriculture

L'élargissement de la méthode du retour par la poste apparaît quelque peu aléatoire dans le cas du recensement de l'agriculture. Il importe d'évaluer soigneusement l'incidence d'une telle mesure sur les coûts, les délais et la couverture potentielle.

4.1 Coûts

Avec la méthode de la reprise, le recensement de l'agriculture entraîne strictement des coûts de dénombrement. Une partie des dépenses est, en fait, imputée au recensement de la population. En effet, l'heure d'arrivée, la durée des déplacements, les frais de déplacement, etc., sont déclarés sous le code de projet du recensement de la population. Ce ne sera plus le cas si la méthode du retour par la poste est élargie au recensement de l'agriculture. L'expérience a révélé que l'intervention du recenseur est plus déterminante lorsqu'il s'agit de remplir le questionnaire agricole que celui sur la population. Les rejets au contrôle et les cas de non-réponse ne seront pas nécessairement les mêmes que pour le recensement de la population. Le manque de renseignements sur des facteurs tels que ce à quoi il faut s'attendre comme taux de réponse par la poste, comme taux de rejet au contrôle, comme succès du suivi téléphonique, comme nombre de suivis sur place nécessaires, etc., fait qu'il est particulièrement difficile d'estimer l'accroissement du coût financier lié à l'adoption de la méthode du retour par la poste. Il faudrait procéder à une étude détaillée pour mesurer cet accroissement. Bien que cette hypothèse ne repose sur rien de ferme, d'après des évaluations antérieures, l'accroissement pourrait être de l'ordre de $500,000.

On devra également prendre en considération le fait que, dans la mesure où le recensement de l'agriculture est effectué en même temps que celui de la population, cela empêche de réduire les coûts liés à ce dernier vu qu'on ne peut augmenter librement la taille des tâches.

4.2 Délais

Les résultats de l'essai portant sur l'élargissement de la méthode du retour par la poste, effectué en 1977, ont indiqué que les délais de la collecte risquaient de poser des problèmes dans les secteurs à forte concentration d'exploitations agricoles, particulièrement s'il y a un nombre important d'exploitants non résidents.

4.3 Couverture

L'essai portant sur l'élargissement de la méthode du retour par la poste, effectué en 1977, a également révélé qu'il y avait un risque que le taux de sous- dénombrement soit plus important dans les secteurs agricoles qui ne sont pas couverts par les plans de township.

Les points que nous venons d'énoncer devront être étudiés en détail si l'on envisage d'appliquer la méthode du retour par la poste au recensement de l'agriculture.

Il conviendrait notamment d'analyser les possibilités suivantes:

- réviser le contenu du questionnaire agricole en le simplifiant afin que celui-ci soit plus facile à remplir, ce qui permettrait d'accroître le taux de réponse par la poste et de réduire le nombre des rejets au contrôle;
- réduire les étapes du contrôle en ce qui a trait aux questionnaires agricoles;
- élaborer un programme publicitaire pour inciter les exploitants agricoles à renvoyer leur questionnaire par la poste;
- étudier les points forts et les points faibles des autres enquêtes relatives à l'agriculture, notamment l'enquête nationale sur les fermes qui utilise la méthode du retour par la poste avec suivi téléphonique;
- remettre aux recenseurs des listes des exploitants non résidents;
- ajouter au questionnaire sur la population une question de sélection du type: "Avez-vous la charge d'une exploitation agricole?".

Il est également suggéré d'envisager la possibilité d'effectuer le recensement de l'agriculture à un autre moment que celui du recensement de la population. On pourrait identifier les exploitants agricoles à l'occasion du recensement de la population et adopter la méthode d'envoi et retour du questionnaire par la poste avec suivi sur place pour les cas de non-réponse et pour les cas n'ayant pu être réglés par téléphone. Cette façon de faire

permettrait, dans le cadre du recensement de la population, de réaliser des économies grâce à une révision des tâches et également, dans le cadre du recensement de l'agriculture, de résoudre les problèmes occasionnés par le fait que recenseur et recensés se connaissent.

5. Autres points à prendre en considération

5.1 Réaction du public

Une étude de la motivation effectuée à la suite de l'essai portant sur l'élargissement de la méthode du retour par la poste, en 1977, a souligné qu'il n'y avait aucune raison pour que les recensés réagissent mal à l'adoption de cette méthode dans les régions rurales. De fait, il y a tout lieu de penser que la réaction sera favorable. Une telle méthode donne aux recensés l'impression que la confidentialité des données est mieux assurée et réduit les risques des interférences entre recenseur et recensés.

5.2 Publicité

La confusion et les inconvénients des "déversements" d'information dus à la publicité sont reconnus, qu'il s'agisse de la méthode du retour par la poste ou de la reprise. Cela est particulièrement vrai dans le cas des recensés demeurant dans des secteurs de reprise qui sont la cible de l'information émanant des centres urbains. L'utilisation d'une méthode standard et, partant, d'un programme publicitaire unique, permettrait d'éviter une telle confusion.

5.3 Normalisation des procédures

L'adoption de l'autodénombrement comme méthode unique permettrait de réduire, voire d'éliminer, certains documents relatifs aux procédures ou servant à la formation et simplifierait le travail des surveillants qui n'auraient plus à superviser le travail d'un personnel qui, dans certains secteurs, applique des méthodes différentes.

5.4 Couverture de la population

Il ne semble pas que l'élargissement de la méthode du retour par la poste risque de nuire à la couverture du recensement. La couverture est déterminée essentiellement à l'occasion de la livraison; or, les procédures de livraison sont les mêmes qu'il s'agisse de la reprise ou du retour par la poste.

Conclusion

En résumé, il est recommandé d'envisager sérieusement l'élargissement de la méthode du retour par la poste pour le recensement de 1991. Il est admis qu'une telle option présente un certain nombre de risques, notamment en ce qui a trait au recensement de l'agriculture; néanmoins, ces risques sont relativement minimes. Les responsables de la Division des opérations régionales sont d'avis que l'adoption de la méthode du retour par la poste comme méthode unique, pour le recensement de 1991, peut se révéler aussi efficace qu'efficiente et, par conséquent, ils appuient ce projet.

TABLE RONDE
LE CADRE DE LA COLLECTE EN 1991 - TROIS SCÉNARIOS

CONFIDENTIALITÉ DES DONNÉES ET ASPECTS JURIDIQUES

RICHARD BARNABÉ

OPÉRATIONS RÉGIONALES - QUÉBEC
STATISTIQUE CANADA

Introduction

La légalité de l'opération et la confidentialité des données sont deux aspects primordiaux de tout recensement. Le présent exposé a pour but de déterminer quelles seront ou quelles pourraient être les dispositions à respecter à ce chapitre et quelles mesures devront être prises pour y faire face aux prochains recensements. Certains des commentaires qui suivent pourront paraître choquants, voire hérétiques, à quelques-uns du fait que les scénarios qui sont décrits s'écartent sensiblement des pratiques établies en ce qui a trait au recensement.

Toutefois, les caractéristiques de telles pratiques, qui sont essentiellement d'ordre technique, pourraient très bien perdre de leur importance face aux impératifs socio-politiques si l'on veut que le recensement reste faisable et acceptable à l'avenir.

Nous aborderons d'abord les aspects juridiques du recensement, puis nous nous replacerons dans le contexte global pour étudier la réaction de l'opinion publique face à ces dispositions légales et, enfin, nous déterminerons les implications qui en résultent pour le Bureau du point de vue de l'établissement des politiques et au niveau opérationnel.

Contexte juridique

Le contexte juridique du recensement est déterminé par les lois pertinentes mais les particularités et la teneur précises en sont modelées par la jurisprudence accumulée. Au Canada, le contexte juridique du recensement est à peu près le même que dans les autres pays de l'Organisation de coopération et de développement économiques (OCDE): une loi, la Loi sur la statistique, autorise Statistique Canada à effectuer le recensement et précise le mandat du Bureau ainsi que les obligations de celui-ci et celles des citoyens. Cependant, l'interprétation de la Loi sur la statistique par les tribunaux est déterminée par l'incidence d'autres lois, notamment la Charte des droits et libertés et la Loi sur la protection des renseignements personnels.

Les réalités juridiques et constitutionnelles imposent certaines limites qui touchent tant la nature que l'étendue des activités de collecte de renseignements par des organismes gouvernementaux. On voit poindre une doctrine appuyée sur la notion de "limites raisonnables" et sur le droit des gens à la protection des renseignements personnels, selon des dosages qui peuvent différer d'un pays à un autre.

En Allemagne et aux Pays-Bas, le recensement a été reporté ou aboli à la suite d'une remise en question devant les tribunaux.

Au Canada, aux États-Unis et au Royaume-Uni, les tribunaux reconnaissent habituellement le bien-fondé et le caractère obligatoire du recensement, encore que certains aspects aient été contestés en Cour.

Au Danemark, la collecte de renseignements directement à partir des dossiers administratifs permanents est autorisée et fortement appuyée par les autorités exécutives aussi bien que judiciaires.

Toutefois, dans l'ensemble, on observe une tendance générale vers un meilleur équilibre entre la satisfaction des besoins légitimes de données des organismes gouvernementaux d'une part et la protection des droits des personnes d'autre part. Une telle orientation est essentielle car, dans nombre de pays, ce domaine particulier est encore un terrain d'essai pour les juristes. En conséquence, il est probable que la jurisprudence accumulée dans chaque pays aura une portée plus universelle que ce n'est habituellement le cas pour d'autres domaines mieux cernés par une forte tradition locale.

Cela veut dire que nous pouvons nous attendre à ce que les tribunaux continuent d'appuyer nos activités dans la mesure où nos méthodes et nos pratiques respectent l'équilibre recherché. En d'autres mots, cela soulève la question de l'identification des données, du stockage des renseignements identifiables et du respect des normes de sécurité dans le cadre de nos activités opérationnelles.

Opinion publique

Les questions juridiques sont fondamentales car elles définissent notre cadre de travail. Néanmoins, le succès de tout recensement dépendra toujours de la reconnaissance de sa nécessité par le public et du degré de coopération de ce dernier au moment de la collecte des renseignements requis.

L'appui que le public apporte au recensement repose sur trois éléments fondamentaux:

- le caractère légal et obligatoire du recensement;
- la reconnaissance de l'utilité du recensement;
- la crédibilité des garanties que nous offrons relativement à la confidentialité des renseignements recueillis et à leur utilisation.

Ces trois éléments permettent de contrer les pressions négatives qui pourraient miner cet appui. On pourrait définir ces pressions négatives comme étant les manifestations d'un quelconque "syndrome orwellien": une certaine inquiétude vis-à-vis de l'habitude contractée par les gouvernements de recueillir des renseignements sur chacun des citoyens afin de mieux gérer les divers aspects de leur existence. L'emprise de ce syndrome est aggravée ou atténuée selon l'opinion que le public a de la façon dont le gouvernement assure le bien-être économique et social de ses citoyens. Néanmoins, en ce qui a trait au recensement, trois préoccupations sont généralement observées, quelle que soit l'opinion qui prévaut:

1. Inquiétude face au volume de renseignements accessibles au gouvernement.
2. Préoccupation en ce qui a trait à l'utilisation faite de ces renseignements.
3. Sentiment que la collecte de renseignements constitue une violation des droits de chacun à la protection des renseignements personnels.

Un recensement ne peut être réussi que si l'on parvient à contrôler adéquatement ces trois types de préoccupation.

Perspectives d'avenir

Avant d'analyser la meilleure manière de faire face à ces diverses inquiétudes, nous allons essayer de déterminer ce que pourrait être le contexte des prochains recensements. À mon avis, il pourrait présenter les caractéristiques suivantes:

- La société à laquelle nous appartenons sera vraisemblablement de plus en plus prompte à contester et le recensement sera plus souvent remis en question devant les tribunaux.
- La confidentialité des renseignements personnels et le droit à la protection de la vie privée demeureront une préoccupation majeure.
- Les progrès technologiques en matière de traitement de l'information iront s'accentuant et les implications (tant positives que négatives) seront plus largement connues.
- L'intérêt accru pour les aspects sociaux de l'existence et pour une plus grande diversité de groupes particuliers augmentera la demande de données détaillées, actuelles et fiables.

Pour le Bureau, ces caractéristiques s'avéreront à la fois une source de possibilités et de problèmes. Il importe que nous sachions dès maintenant comment nous ferons pour y répondre le moment venu.

Conclusion

Nous devons être prêts à fournir une réponse adéquate sur les plans suivants:

1. Confidentialité

Nous devons fournir des garanties valables en ce qui a trait à trois types de menaces spécifiques:

Menace directe:

Risque qu'un employé divulgue, volontairement ou maladroitement, des renseignements confidentiels. Il s'agit du type de menace auquel il nous est le plus facile de faire face car nous disposons des mécanismes judiciaires nécessaires. Nous devons toutefois avoir soin de ne pas sacrifier à la prudence dans le seul but d'atteindre une plus grande efficacité ou une plus grande efficience. Il convient de ne pas perdre de vue que les réductions de coûts et la recherche d'une plus grande rapidité d'exécution peuvent s'avérer très préjudiciables à ce chapitre.

Menace indirecte:

Danger lié à l'existence de fichiers de renseignements distincts, parfaitement protégés,

mais dont la fusion ou l'appariement pourrait entraîner la divulgation de renseignements confidentiels. Des techniques particulières, telles que l'arrondissement aléatoire, l'application d'algorithmes de suppression et la restriction de l'accessibilité des données permettent de réduire ces risques. Nous devons, cependant, demeurer extrêmement vigilants.

Menace perçue:

Question de l'utilisation qui sera faite, immédiatement et ultérieurement, des renseignements recueillis par des autorités gouvernementales mal intentionnées (de l'avis du public). Nous n'avons guère de moyens de réduire cette menace aussi longtemps que les renseignements que nous recueillons et stockons sont identifiables. Le système adopté par le Bureau s'est avéré tout à fait adéquat jusqu'à présent, mais il pourrait ne plus en être de même à l'avenir étant donné l'évolution de l'opinion à ce chapitre.

2. Identification

Dans la mesure où la collecte et le stockage de renseignements identifiables demeurent une source majeure de préoccupation en dépit des garanties qui sont données relativement à la confidentialité des données, il faut envisager de prendre des mesures plus sévères. Le fait de ne pas recueillir le nom des recensés ou de détruire toute donnée d'identification au début des opérations de recensement aurait une incidence considérable sur des concepts tels que la formation des familles, sur la conception du questionnaire, sur les techniques de suivi et sur l'évaluation de la couverture. Pour rébarbatives que de telles solutions puissent être, nous devons les évaluer et envisager des alternatives possibles, ne serait-ce que pour être prêts à toute éventualité advenant que des pressions extérieures non encore définies nous poussent dans cette direction. Il s'agit avant toute chose d'éviter d'être acculés, sans préparation, à une telle issue.

3. Aspects juridiques

Les organismes statistiques devront-ils recommander ouvertement le renforcement des lois relatives au traitement de l'information afin de rassurer le public et de le convaincre qu'il peut coopérer sans risque? Par exemple, la Loi pourrait limiter plus rigoureusement le partage de renseignements, la fusion de fichiers, ou encore la collecte, ou du moins le stockage, de renseignements identifiables. De telles restrictions nous forceraient à renoncer à utiliser certaines possibilités techniques; par ailleurs, ce changement d'attitude pourrait devenir une condition essentielle de la poursuite de nos activités.

Nous pourrions aussi envisager d'opter pour des méthodes de travail totalement différentes comme, par exemple, constituer des registres permanents à partir des fichiers existants (et non facilement controversés) et renoncer à la collecte directe des données. Là encore, les implications légales et politiques de telles solutions sont considérables.

La question fondamentale à laquelle nous devons répondre dans l'immédiat consiste à savoir si nous sommes prêts à adopter une attitude proactive lorsqu'il s'agit de modifier notre façon de faire.

4. Communications

Les organismes chargés d'effectuer les recensements sont très préoccupés de faire de chaque recensement une enquête nationale rentable. Malheureusement, nos efforts dans ce sens et partant, notre aptitude à nous gagner l'appui du public, sont très sporadiques et tendent à se manifester essentiellement en période censitaire. Il conviendrait de repenser cette façon de faire et d'investir davantage de ressources dans ce domaine, même si nous connaissons une période de restrictions, en renonçant à certaines améliorations techniques, améliorations qui nous sont particulièrement chères d'un point de vue professionnel mais qui sont sans doute moins essentielles si l'on se place du point de vue de nos intérets stratégiques à long terme ou du point de vue du public que nous servons.

POINTS IMPORTANTS RELATIVEMENT AUX PERSONNES ÂGÉES ET AUX IMMIGRANTS, QUI PRÉSENTENT UN INTÉRÊT POUR LES OPÉRATIONS DE COLLECTE

PRÉSENTÉ PAR: J. BOYD UNDERHAY

TERRE-NEUVE ET LABRADOR
OPÉRATIONS RÉGIONALES
STATISTIQUE CANADA

Les personnes âgées

Les personnes âgées ont tendance à être plus difficiles à dénombrer pour diverses raisons, telles que:

- leur méfiance générale envers les représentants du gouvernement;
- une certaine confusion face aux formules gouvernementales;
- la crainte de divulguer des renseignements personnels.

En 1991, les personnes âgées seront un groupe puissant et important. En effet:

- elles représenteront 11.8% de la population;
- elles connaîtront un taux de croissance plus rapide que celui des jeunes;
- elles vivront plus longtemps;
- elles seront plus instruites;
- elles dépendront moins des paiements de transfert gouvernementaux;
- elles auront tendance à vivre dans des ménages privés plutôt qu'en établissements;
- elles auront tendance à vivre dans de petits centres urbains.

Nous allons nous arrêter à deux phénomènes qui présentent un intérêt particulier pour le dénombrement des personnes âgées en 1991.

1. La retraite anticipée:

- Il est possible que la tendance à prendre une retraite anticipée, constatée au cours des années 1980, se poursuive pendant les années 1990.
- Si l'on ajoute au nombre de personnes âgées (65 ans et plus) les personnes de 55 à 64 ans (c.-à-d. celles qui ont pris une retraite anticipée), le groupe obtenu représente environ 20.8% de la population totale.
- Comme "population qui présente un intérêt particulier" potentielle, ce groupe aurait la même importance que celui des jeunes "de passage" au cours des années 1980.

2. Les personnes âgées comptent une proportion importante d'immigrants:

- En 1981, les immigrants constituaient près du tiers de la population âgée.
- Si cette portion augmente en 1991, il faudra tenir compte de deux caractéristiques de cette population au moment de la planification de la collecte:

  a) tendance chez les immigrants âgés à conserver leur langue maternelle,
  b) concentration de cette population dans les constructions en hauteur des grandes régions métropolitaines, particulièrement en Ontario, en Colombie-Britannique, au Québec et en Alberta.

La population des immigrants

La population des immigrants dans son ensemble a tendance à être plus difficile à dénombrer. Les facteurs qui compliquent le dénombrement sont les mêmes que dans le cas des personnes âgées mais, en plus, il peut exister un problème de langue.

Les caractéristiques suivantes de la population des immigrants présenteront un intérêt particulier pour le recensement de 1991:

- augmentation possible de la population des immigrants;
- tendance à être plus instruits que la population dans son ensemble;
- problèmes de langue (en 1981, une faible minorité d'immigrants ne pouvaient parler aucune des deux langues officielles);
- tendance à avoir une plus forte proportion de revenus de placements;
- tendance à dépendre davantage de pensions privées;
- tendance à habiter dans des constructions en hauteur;
- tendance à vivre dans des logements multifamiliaux.

Il y a aussi eu une augmentation sensible du nombre d'immigrants d'origine asiatique au cours des dernières années. Ce groupe formera éventuellement une partie de la population des personnes âgées. Ce fait se reflétera dans les caractéristiques sociales de la population des personnes âgées des années 1990.

AUTOCHTONES ET RÉGIONS ÉLOIGNÉES

C. JERRY PAGE

OPÉRATIONS RÉGIONALES - ALBERTA, NORD DE LA SASKATCHEWAN ET T. N.-O.
STATISTIQUE CANADA

Introduction

Les régions éloignées, au même titre que les autochtones, sont un sujet d'inquiétudes pour les agents recenseurs. Les autochtones vivent surtout dans des régions éloignées et la population de ces régions est composée essentiellement d'autochtones.

Régions éloignées

Au Canada, les régions éloignées se trouvent dans les Territoires du Nord-Ouest (T. N.-O.), au Yukon, sur la côte du Labrador et dans la partie septentrionale de la plupart des provinces.

Le présent exposé porte sur les T. N.-O., qui représentent un tiers de la superficie du Canada. Leur population est très faible, seulement 46,000 personnes dont 16,000 Inuit, 9,000 Indiens inscrits, 5,000 Métis et Indiens non inscrits et 16,000 personnes d'origine européenne. On y compte 63 collectivités, dont six regroupent 48.1% de la population totale. Cinquante-sept de ces collectivités ont une population moyenne d'environ 400 personnes. La langue prédominante est l'anglais, mais l'Inuktitut (Inuit), le français et les langues slaves sont aussi parlées.

Les communications et le transport aérien sont satisfaisants, sauf dans le cas des très petites agglomérations.

Deux méthodes de dénombrement ont été utilisées lors du recensement de 1981: le dénombrement ordinaire le 3 juin et un dénombrement anticipé qui a commencé en mars. Le dénombrement anticipé a été effectué pour réduire les coûts et pour assurer la qualité des données. Si nous avions attendu le dégel du printemps (peu de temps après Pâques), de 15% à 20% des autochtones n'auraient pu être dénombrés car ils auraient alors quitté les collectivités qu'ils habitent pour poursuivre leur style de vie traditionnel en pleine nature. On aurait alors abouti à une augmentation importante du sous-dénombrement de la population et des coûts en raison de la difficulté de retrouver cette population éparpillée.

Les agents recenseurs font encore face à de nombreux problèmes comme la difficulté de recruter des autochtones qualifiés qui acceptent de travailler pour une si courte période, le roulement des employés de même que la difficulté d'assurer la surveillance du personnel et de contrôler les tâches terminées dans les collectivités isolées.

Une proposition avait été faite pour le recensement de 1986 selon laquelle une équipe d'agents recenseurs provenant du Sud aurait effectué le dénombrement anticipé dans le Nord. Cela suppose qu'un groupe de personnes arrive par avion dans une collectivité isolée pour y effectuer le dénombrement dans un délai fixé avec l'aide d'agents recenseurs locaux et (ou) d'interprètes, le cas échéant.

Nos objectifs pour le recensement de 1991 sont de réduire l'importance du dénombrement anticipé dans les Territoires du Nord-Ouest et d'améliorer la qualité des données. Pour atteindre ces objectifs, nous devons nous pencher sur plusieurs questions qui sont présentées ci-après:

- Les habitants du Nord désirent être traités de la même façon que les personnes qui habitent le Sud du Canada. Ils ne pensent pas que ce soit équitable de dénombrer toute la population des T. N.-O. au moyen du questionnaire complet (formule 2B). Ils veulent être dénombrés par autodénombrement et par échantillonnage (un ménage sur cinq) comme c'est le cas pour les habitants du Sud du pays.

- Les autochtones désirent pouvoir obtenir des questionnaires dans leur langue maternelle, pas seulement en avoir la traduction. Nous aurons encore le problème de la population autochtone éparpillée en pleine nature. Il ne semble pas y avoir de méthodes qui puissent remplacer le dénombrement anticipé. Cependant, ce type de dénombrement lui-même pourrait poser un problème. Il n'est pas évident que nous puissions légalement effectuer un recensement anticipé et exiger la participation de la population visée.

Les progrès dans les domaines de la technologie, du transport et des communications continueront à améliorer le dénombrement dans les secteurs de recensement par interview et aussi à accroître l'actualité ainsi que la qualité des données.

Autochtones

Les 491,460 personnes qui se sont déclarées autochtones lors du recensement de 1981 représentaient seulement 2% de la population totale. Ces personnes étaient réparties de façon très peu uniforme dans tout le Canada. Près de 60% de la population des T. N.-O. est autochtone, et environ 20% des habitants du Yukon se sont déclarés autochtones. Dans les provinces, le Manitoba ainsi que la Saskatchewan comptent la proportion la plus élevée d'autochtones, plus de 6%. En Alberta et en Colombie-Britannique, 3% des habitants ont dit être d'origine autochtone. (Dans l'Est, une personne sur 100 a été identifiée comme autochtone.) La majorité des autochtones vivaient dans l'Ouest.

Les Indiens inscrits ne vivent pas tous dans des réserves. Les réserves indiennes sont des territoires réservés à des fins spéciales à la suite de traités conclus entre le gouvernement fédéral et les Indiens. Moins de 60% des Indiens inscrits résident dans des réserves. Quatre autochtones sur 10 vivent dans des régions urbaines alors que la proportion est d'environ huit sur 10 dans le cas des autres Canadiens. Nous avons relevé des différences importantes entre les groupes autochtones. Deux Inuit sur 10 et trois Indiens inscrits sur 10 vivaient dans des régions urbaines comparativement à six Métis sur 10 et à sept Indiens non inscrits sur 10. L'urbanisation des autochtones pourrait cependant augmenter à la suite de la migration vers les régions urbaines des jeunes autochtones à la recherche d'emploi.

La population autochtone est très jeune; l'âge moyen est de 23 ans comparativement à 32 pour les non-autochtones. À l'exception des Inuit, l'anglais est la langue parlée le plus souvent à la maison par les autochtones. Les autochtones ont un niveau de scolarité moins élevé que les autres Canadiens; seulement 25% des autochtones sont titulaires d'un diplôme d'études secondaires comparativement à 50% des autres Canadiens. Par rapport aux autres Canadiens, moins d'autochtones sont employés et (ou) font partie de la population active. Ces derniers ont tendance à avoir des emplois saisonniers et, à cause de leur niveau de scolarité relativement faible, on ne les retrouve pas dans les emplois bien payés. À mesure que le niveau de scolarité continuera d'augmenter, les autochtones obtiendront de meilleurs emplois dans l'avenir, en dépit de la distance qui sépare leur domicile des principaux marchés du travail.

Les revendications territoriales et la campagne pour obtenir l'autonomie pourraient augmenter la résistance au gouvernement fédéral dans son ensemble et au recensement en particulier. Les autochtones prétendent, en général, qu'on ne les consulte pas à propos des procédures de dénombrement et qu'ils n'obtiennent aucun avantage direct du recensement, car les subventions proportionnelles au nombre d'habitants sont accordées aux municipalités et non à la bande. Comme les municipalités ne fournissent généralement aucun service aux réserves, ces dernières n'y voient aucun avantage.

Par contre, si nous pouvons démontrer la valeur et l'importance de bonnes données statistiques, il se peut que les autochtones soient plus réceptifs à nos efforts de collecte des données.

Tableau 1: Répartition de la population autochtone en 1981, Canada

TOTAL DE LA POPULATION AUTOCHTONE	491,460
INUIT	25,390
INDIENS INSCRITS	292,700
INDIENS NON INSCRITS	75,110
MÉTIS	98,260

Résumé de la période de questions

Après l'exposé de **M. Kazmaier**, *les participants ont soulevé quelques questions reliées au suivi téléphonique.*

Préoccupation des recensés concernant la protection de la vie privée

Tout d'abord, il s'est dégagée une appréhension générale en regard de la réaction des recensés face au suivi téléphonique, particulièrement dans les cas où le recensé n'a pas fourni son numéro de téléphone. Ce problème s'accentuera dans l'avenir à mesure que l'inquiétude du public relativement à la confidentialité s'accroîtra par suite de l'augmentation du volume de renseignements personnels emmagasinés sur support informatique. Le recensement est d'autant plus vulnérable de ce point de vue que son succès dépend de la collaboration du public. À date les recensements des États-Unis et du Canada n'ont pas vraiment subi l'effet de ce problème potentiel.

Incidence de la nouvelle technologie

Un autre sujet d'appréhension est l'incidence des divers progrès technologiques sur l'utilisation du téléphone. Le personnel du recensement est déjà aux prises avec le problème de l'usage croissant des répondeurs téléphoniques. Le United States Bureau of the Census (USBC), lors des expériences de suivi téléphonique centralisé qu'il a effectuées, a établi que la façon la plus pratique de traiter ces cas serait de laisser un message et de répéter les appels. Malheureusement, les expériences dans ce domaine en sont encore à leurs débuts.

Un progrès technologique plus positif rend possible les recoupements entre les numéros de téléphone, les codes d'adresses géographiques, et enfin l'information. Ce progrès présente beaucoup de possibilités pour l'avenir. Une application immédiate d'un tel mécanisme pourrait permettre aux téléphonistes des Services auxiliaires téléphoniques d'apparier, au moment de la collecte sur le terrain, une adresse à son code géographique quand les recensés appellent pour signaler qu'ils n'ont pas encore reçu de questionnaire.

Utilisation accrue du suivi téléphonique

Le troisième problème soulevé se rapporte à l'utilisation accrue du suivi téléphonique pour résoudre les cas de questionnaires rejetés au contrôle et des ménages non répondants. Le recensement d'essai de 1986, qui sera effectué par le USBC, s'attaquera à ce problème. L'évaluation des résultats de l'essai fournira des données sur les implications relativement à la couverture et à la qualité des données ainsi qu'une idée des conséquences au chapitre des coûts.

L'exposé de **M. Hicks** *a mis en évidence deux aspects essentiels de l'élargissement de la méthode du retour par la poste des secteurs de reprise actuels.*

Le premier aspect est l'incidence possible d'un tel changement sur le recensement de l'agriculture. L'élargissement de la méthode du retour par la poste pourrait, en effet, avoir un effet néfaste sur la couverture des exploitations agricoles dans le cadre du recensement de l'agriculture. Les raisons en sont que la majorité de ces exploitations étaient jusqu'à présent dénombrées au moyen de la reprise et que l'intervention du recenseur est un facteur déterminant de l'intégralité de la couverture.

Aussi, l'élargissement de la méthode du retour par la poste aux secteurs de reprise actuels implique une révision des limites des secteurs de dénombrement puisque, dans les secteurs de retour par la poste, chaque recenseur peut dénombrer un nombre plus important de ménages. Toutefois, l'utilisation accrue de l'ordinateur pour produire le matériel géographique nécessaire pour la collecte pourrait atténuer l'effet de la révision des limites des secteurs sur la charge de travail globale.

SÉANCE: GÉOGRAPHIE DU RECENSEMENT

Président: D. Ross Bradley
Division de la Géographie
Statistique Canada

Mardi, le 8 octobre 1985

SOUTIEN GÉOGRAPHIQUE POUR LE RECENSEMENT DÉCENNAL DE 1990

SILLA G. TOMASI

GEOGRAPHY DIVISION
U.S. BUREAU OF THE CENSUS

Introduction

La mission du U.S. Bureau of the Census est de fournir à notre Congrès, à notre pouvoir exécutif ainsi qu'au public dans son ensemble des statistiques de base relatives à la population et à l'économie de notre nation. Le succès d'un recensement ne repose pas seulement sur l'efficacité de la collecte des données, mais aussi sur la manière dont ces données sont liées à des régions géographiques. Comme les pièces d'un gigantesque casse-tête, des milliers de petites unités géographiques d'ordre administratif, employées pour contrôler le déroulement du recensement, doivent s'adapter les unes aux autres et inclure chaque mille carré du pays, et chaque mille carré doit être inclus une fois et une seule. Les millions d'éléments de données distincts doivent alors être transformés en unités statistiques d'utilisation commode qui sont groupées et regroupées en de nombreuses combinaisons selon les régions géographiques conçues pour répondre à des intérêts divers. Sans possibilité d'affecter ou de relier les données à des régions particulières, les données n'ont que très peu de valeur pour les utilisateurs, mis à part les totaux nationaux; les données sont intéressantes, mais leur utilité est grandement limitée. Les données du recensement répondent à la plupart des besoins des utilisateurs seulement lorsqu'elles sont reliées à une région particulière, ou à une combinaison de régions, sur la surface de la terre; c'est donc le cadre géographique du recensement qui donne leur signification aux données en permettant d'identifier les unités géographiques par leur nom, par leurs limites et par leurs liens réciproques (voir le tableau 1).

Fonction de soutien géographique

Il n'est donc pas surprenant d'apprendre que la fonction de soutien géographique est un des domaines les plus importants qui sont approfondis au cours de la planification pour le recensement

Tableau 1. Types de régions géographiques - Recensement de 1980

RÉGIONS POLITIQUES	RÉGIONS STATISTIQUES
États-Unis	Régions (4)
États et équivalents (57)	Divisions (9)
États (50)	Régions statistiques unifiées
D.C. (1)	normalisées - SCSA (17)
Régions éloignées (6)	Régions statistiques métropolitaines
Comtés, paroisses et autres	normalisées - SMSA (323)
équivalents (3,231)	Régions urbanisées - UA (373)
Divisions civiles - MCD (30,491)	Divisions de comtés de recensement - CCD (5,512)
Localités constituées (19,176)	Territoires non organisés (274)
Réserves indiennes américaines (275)	Localités désignées pour le recensement - CDP (3,733)
Régions constituées de sous-réserves indiennes (228)	Secteurs de recensement (43,383)
Villages autochtones d'Alaska (209)	Secteurs de numérotation d'îlots - BNA (3,404)
Districts de représentation au Congrès - CD (435)	Secteurs de dénombrement - ED (102,235)
Circonscriptions électorales (36,361)	Groupes d'îlots - BG (156,163)
(réparties dans 23 états)	(parties totalisées - 200,043)
Districts scolaires (16,075)	Îlots (2,473,679)
Banlieues (28,381)	(parties totalisées - 2,545,416)
Codes postaux (environ 37,000)	Zones analytiques de circulation (environ 160,000)

décennal de 1990. Les opérations géographiques interviennent dans presque toutes les phases des travaux du recensement, ou encore ont une incidence sur celles-ci ou sont déterminées par elles; leur rôle est primordial à toutes les étapes de la planification, du dénombrement et de la totalisation. L'identification des unités géographiques et de leurs limites est la base du contrôle administratif lié au dénombrement ainsi qu'au dépouillement et à la totalisation des résultats.

Fondamentalement, tous les aspects de la fonction de soutien géographique (et cartographique) sont centrés autour de deux tâches principales: tout d'abord, l'affectation de chaque adresse domiciliaire à son bon emplacement géographique et, deuxièmement, la classification de chaque emplacement selon les diverses régions de totalisation représentées dans le recensement. Les mécanismes utilisés pour atteindre ces buts doivent prévoir qu'il faut fournir trois outils géographiques importants: les cartes, les fichiers de référence des adresses, appelés fichiers GBF/DIME en 1980, et un fichier de référence géographique, appelé le Master Reference File en 1980.

Cartes

Les cartes sont une représentation graphique de la terre. Les cartes du recensement montrent les rues, les voies ferrées, les cours d'eau et les autres types de caractéristiques qu'un recenseur peut repérer sur le terrain lorsqu'il effectue la collecte des données dans sa zone d'affectation. Les cartes montrent aussi les polygones, que nous appelons îlots, constitués par ces caractéristiques auxquels nous attribuons un code numérique unique afin de représenter les entités géographiques qui présentent un intérêt pour le recensement. Une fois les limites des zones d'affectation sur le terrain indiquées sur une carte du recensement, le recenseur peut utiliser cette carte pour faire le tour d'un îlot à pied ou en automobile, pour inscrire dans un registre de listage les adresses qui se trouvent de chaque côté de l'îlot et le numéro de l'îlot correspondant. Ce faisant, le recenseur a "géocodé" l'adresse; c'est-à-dire qu'il l'a attribuée à un emplacement géographique, ce qui correspond à la première des deux fonctions de soutien géographique mentionnées plus haut.

Fichiers de référence des adresses

Alors qu'une carte décrit la terre sous forme géographique à un humain, les fichiers de référence des adresses (ou fichiers GBF/DIME, qui ressemblent à vos fichiers principaux de région (FPR)) décrivent la terre, sous forme graphique, à un ordinateur. C'est-à-dire que l'ordinateur peut effectuer les tâches de géocodage, à la place du recenseur, pour les régions du pays où cela est possible. Afin d'utiliser un ordinateur pour attribuer une adresse à son emplacement géographique, il est nécessaire de saisir les renseignements géographiques indiqués sur une carte de recensement traditionnelle, c'est-à-dire les rues, les voies ferrées, les cours d'eau, les identificateurs d'îlots, etc., et d'ajouter des renseignements sur les tranches d'adresses qui se trouvent de chaque côté d'une rue entre deux caractéristiques, indiquées sur la carte, qui forment une intersection avec cette rue. Le fait d'ajouter des renseignements sur les tranches d'adresses permet à l'ordinateur de "voir" les adresses qui font partie de chaque îlot au moyen d'algorithmes d'appariement informatique qui effectuent le travail de géocodage fait auparavant par un recenseur. Comme les fichiers automatisés (fichiers de référence des adresses) contiennent uniquement des renseignements géographiques et non des renseignements sur la longueur des rues ou leur position sur la terre, il faut utiliser une carte quand le géocodage est effectué manuellement. Pour le recensement de 1980, quelque 55 millions de ménages, soit environ 60% de l'ensemble des logements dénombrés lors du recensement, se trouvaient dans des régions où l'on pouvait utiliser les fichiers de référence des adresses pour le géocodage.

Fichier de référence géographique

Les deux outils géographiques décrits jusqu'ici sont surtout destinés à être utilisés pour la première des fonctions de soutien géographique, celle qui consiste à attribuer une adresse à un emplacement géographique. La seconde fonction de soutien, la classification de chaque emplacement géographique selon les unités de totalisation reconnues dans le recensement, est effectuée au moyen d'un fichier de référence géographique, que nous avons appelé le Master Reference File, ou MRF, pour le recensement décennal de 1980. Ce fichier montre, sous une forme assimilable par l'ordinateur, les liens entre les unités géographiques qui figurent sur l'ensemble des cartes du recensement qui couvrent tous les États-Unis et leurs possessions. Au moyen de ce fichier, nous pouvons effectuer le lien entre l'emplacement géographique auquel une adresse a été attribuée (le numéro d'îlot consigné par le recenseur sur le terrain ou attribué par l'ordinateur) et toutes les unités géographiques de niveau supérieur pour lesquelles les données seront totalisées.

L'expérience de 1980

Les trois produits de soutien géographique, les cartes, les fichiers de référence des adresses et le fichier de référence géographique, possèdent

plusieurs points communs; en d'autres mots, il s'agit de trois façons différentes de décrire une partie de la surface de la terre. Des problèmes avec ces produits géographiques au cours du recensement de 1980 ont causé de la confusion chez les employés du Bureau of the Census travaillant sur le terrain et chez les utilisateurs des données. Ces problèmes sont dus au fait que ces trois outils géographiques ont été élaborés lors d'opérations manuelles distinctes et complexes auxquelles ont travaillé des centaines de commis, dont un grand nombre venaient d'être embauchés. Des erreurs différentes ont été faites pour chaque produit, ce qui a entraîné des incohérences entre les produits finals. Il est inévitable que des erreurs se produisent dans une opération au cours de laquelle des millions de régions et d'identificateurs géographiques ont été inscrits à la main sur des cartes et deux fichiers informatiques indépendants élaborés manuellement. Comme cela était prévu, nous n'avons pas pu assurer une cohérence parfaite entre ces trois produits géographiques de base. La très grande majorité des identificateurs concordaient parfaitement d'un produit à un autre, mais où il existait des incohérences, cela a causé des problèmes lors de la collecte, de la totalisation et de l'utilisation des données. De plus, les processus adoptés par le Bureau of the Census pour élaborer ces produits géographiques utilisant beaucoup de main-d'oeuvre, les délais de production ont été très longs. Les opérations complexes et fonctionnellement distinctes utilisées pour élaborer les produits géographiques de 1980, tout comme les processus utilisés pour les recensements antérieurs, n'avaient pas été conçus en fonction de l'ère de l'informatique.

Dans une tentative pour surmonter les problèmes liés au fait que les outils géographiques étaient produits manuellement et individuellement, la Geography Division a mis sur pied un programme destiné à produire un système de soutien géographique (Geographic Support System) unifié. Nous avons conçu et sommes en train de mettre au point un tel système pour le recensement de 1990. Ce système permettra d'éviter un bon nombre des problèmes que nous avons connus dans le passé. Notre objectif est donc d'enregistrer tous les renseignements géographiques pertinents relatifs à une région dans un **seul fichier informatique.** Ce fichier permettra à l'ordinateur de produire des cartes pour les opérations sur le terrain et pour les publications, et d'effectuer des opérations de totalisation des données pour toute entité géographique dont les limites ont été enregistrées dans le fichier. Nous appelons ce fichier, ainsi que le logiciel informatique connexe, le Topologically Integrated Geographic Encoding and Referencing System ou encore "TIGER" (voir la figure 1). Le fichier TIGER se trouve au coeur du système.

Il n'entre pas dans le cadre du présent document d'exposer en détail l'élaboration du système TIGER pas plus que d'expliquer la structure du fichier TIGER. Il semble toutefois approprié de présenter un bref résumé de ces deux aspects. Quiconque est intéressé à obtenir plus de renseignements est prié de me le faire savoir ou d'écrire au Chief, Geography Division, Bureau of the Census, Washington, D.C. 20233. À mesure que nous disposons des spécifications du fichier TIGER, nous en fournissons des exemplaires à la Sous-division de la géocartographie, Division des services et du développement informatiques, de Statistique Canada.

Le système TIGER

Le concept fondamental du fichier TIGER est la création d'un système de base de données géographiques unique fondé sur des cartes numériques ou assimilables par l'ordinateur. Ce fichier doit couvrir tous les États-Unis et toutes les autres régions pour lesquelles le Bureau of the Census effectue un recensement. Pour créer la structure interne du fichier TIGER nous avons appliqué les concepts de la topologie, branche des mathématiques qui décrit les liens spatiaux entre points, lignes et surfaces dans un plan à deux dimensions. La construction du fichier au moyen de ces principes nous permet de fabriquer une base de données informatique élégante, pouvant effectuer beaucoup d'autovérifications et contenant toutes les caractéristiques de base des cartes ainsi que les limites, noms, codes géographiques et tranches d'adresses des diverses entités géographiques. Nous mémorisons les renseignements géographiques dans le fichier TIGER au moyen de listes et de répertoires entre lesquels le couplage et les renvois sont effectués automatiquement. Les segments de droite, les points, et les codes de classification des régions géographiques sont tous reliés de sorte que si l'un quelconque d'entre eux subit des modifications, les éléments et les relations connexes sont aussi modifiés (Marx, 1984). Seront aussi intégrés à cette **unique** base de données de cartes numériques tous les renseignements nécessaires pour permettre (au moyen de programmes d'application) la production des cartes du recensement à l'aide de traceurs commandés par ordinateur, l'affectation des questionnaires du recensement aux emplacements géographiques et le catalogage ainsi que la définition relationnelle des régions. Cela signifie qu'il y aura un accord parfait entre la cartographie et le dépouillement géographique et que le problème des incohérences entre les produits géographiques sera réglé. Le système TIGER nous permettra aussi de produire des cartes dont la qualité sera uniformément élevée, ce qui n'était pas toujours possible auparavant.

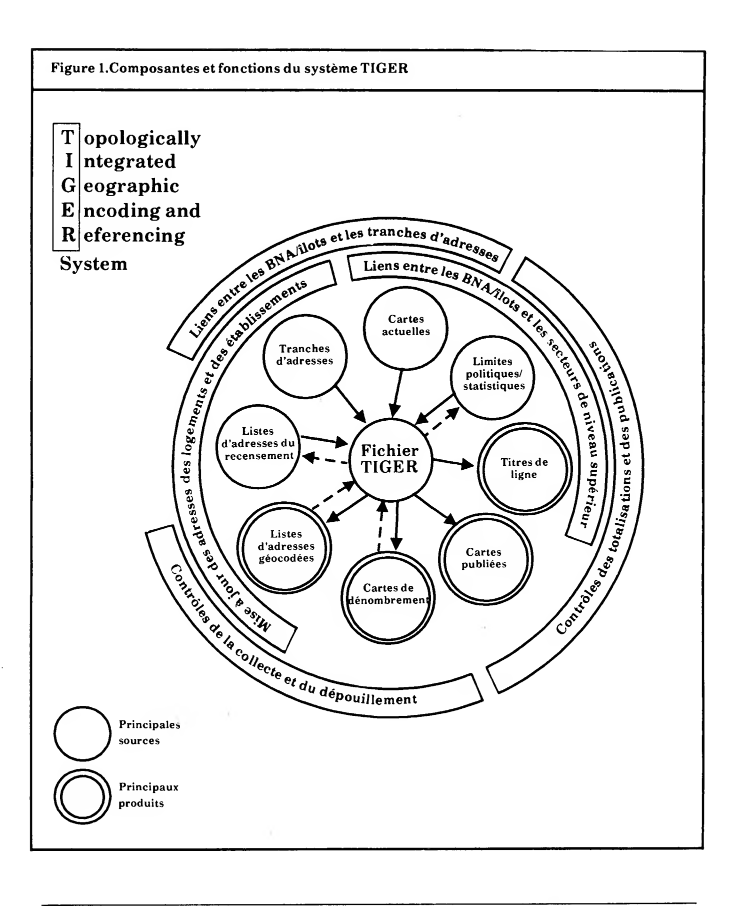


Figure 1.Composantes et fonctions du système TIGER

Le fichier TIGER comprendra un ensemble d'articles représentant la position des routes, des cours d'eau, des voies ferrées, des limites politiques et statistiques et d'autres caractéristiques nécessaires au recensement ainsi que leurs attributs, tels que le nom, le code géographique, les tranches d'adresses, la catégorie de caractéristique et ainsi de suite (Broome, 1984).

Chaque article contiendra un certain nombre de renseignements, tels que:

(1) le nom et le genre de caractéristique; c'est-à-dire si la caractéristique est une route ou une rue, un passage d'eau, une voie ferrée, une limite politique, etc.;

(2) les coordonnées qui définissent les points d'intersection le long de la caractéristique, avec les autres traits distinctifs de celle-ci; par exemple, les vecteurs de description de courbe qui définissent la forme de la caractéristique;

(3) les tranches d'adresses situées entre des points d'intersection pour les articles se rapportant à des rues ou des routes, en plus du nom du bureau de poste et du code postal ("ZIP code") de chaque tranche d'adresses;

(4) les codes de la (des) région(s) géographique(s) qui s'appliquent à chaque segment de la caractéristique d'après le lien géométrique avec la caractéristique des limites dans le fichier pour chaque entité géographique; et

(5) d'autres particularités associées à l'article; par exemple, les principaux centres d'emploi ou les structures résidentielles situées le long de la caractéristique.

Une fois construites, les première, deuxième et troisième caractéristiques de l'article du fichier TIGER fournissent un cadre géographique à partir duquel il est possible de créer les cartes nécessaires aux travaux sur le terrain ou à la production des documents; les première, troisième et cinquième caractéristiques permettent d'attribuer les adresses à des régions géographiques particulières; les première, troisième et quatrième caractéristiques constituent une base sur laquelle on peut établir un système de vérification et de contrôle automatisé des questionnaires, ce qui permet la création de tâches de suivi d'après l'adresse de la construction ou le numéro d'ordre dans une perspective géographique; et la quatrième caractéristique fournit le fichier de base à partir duquel on peut obtenir les titres de colonnes des tableaux géographiques et les produits cartographiques sommaires, à des fins de totalisation.

Notre responsabilité est de terminer l'élaboration du fichier TIGER et des éléments connexes du système TIGER conformément au calendrier des activités du recensement décennal de 1990; la mise en exploitation initiale est prévue pour le premier trimestre de l'année civile 1988. Nous considérons que le travail à accomplir pour atteindre nos objectifs peut se diviser en trois grandes étapes, dont l'organigramme de la figure 2 donne une vue d'ensemble.

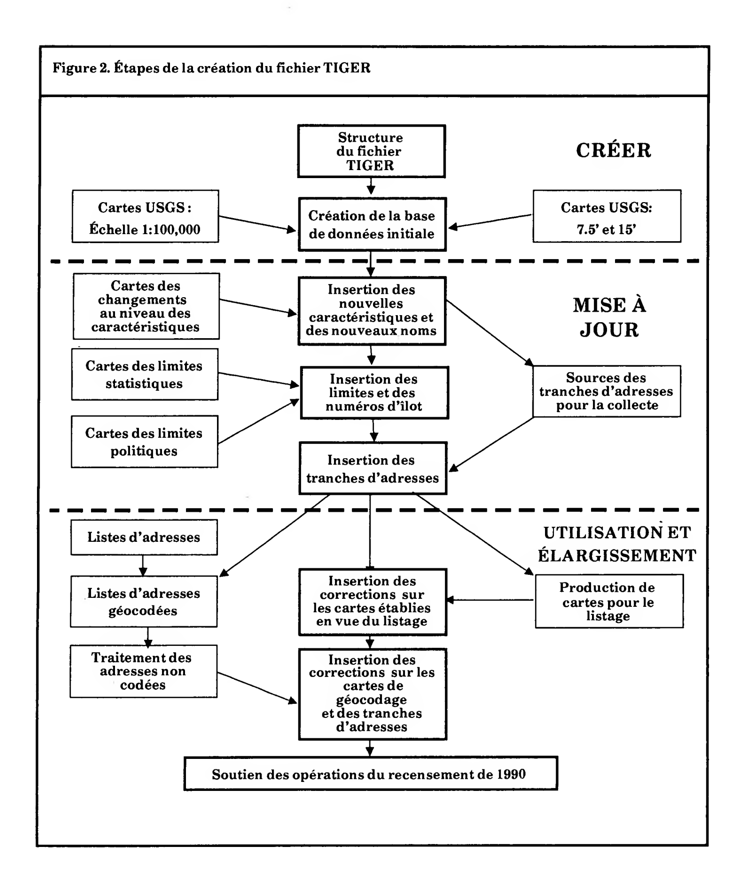
- Tout d'abord, nous devons **créer** la base de données initiale des cartes numériques, la base de cartes sous-jacente sous une forme assimilable par l'ordinateur afin que ce dernier contienne tous les cours d'eau, rues, voies ferrées, etc. Nous travaillons en étroite collaboration avec notre organisme frère, le United States Geological Survey (USGS), pour atteindre cet objectif.

- Deuxièmement, nous devons **mettre à jour** cette base de données de cartes numériques au moyen d'autres renseignements qui permettent de rendre la carte utile aux travaux du Bureau of the Census: il nous faut ajouter des rues dans les secteurs nouvellement construits; il faut introduire les noms de tous les cours d'eau, rues et voies ferrées de sorte que les employés du Bureau of the Census qui travaillent sur le terrain ainsi que les utilisateurs des données puissent s'orienter; il faut insérer les tranches d'adresses associées à chaque segment de rue dans les principales régions urbaines et ajouter les limites et les codes de toutes les régions politiques et statistiques pour lesquelles le Bureau of the Census totalise des données.

- Finalement, nous devons **utiliser** le fichier et faire les **ajouts** nécessaires d'après les résultats des premières applications du fichier. Ici, les activités relatives à la constitution du fichier et à son utilisation sont étroitement liées parce que les opérations qui utilisent le fichier apportent des renseignements valables provenant de l'observation directe faite par les employés sur le terrain afin que la base de données des cartes numériques corresponde à la réalité.

Évidemment, le processus de constitution du fichier TIGER comprend de nombreuses autres étapes auxquelles s'ajoutent les étapes propres à sa fonction de soutien du recensement de 1990. Pour mener à bien un travail de cette importance dans une période aussi courte, il faudra dépenser des sommes considérables et y employer, littéralement, des centaines de géographes, de cartographes, de programmeurs et de commis.

Figure 2. Étapes de la création du fichier TIGER



Conclusion

Le concept du système TIGER n'a pas été conçu du jour au lendemain. Il résulte du travail ardu et du dévouement d'un petit groupe de personnes qui, en 1981, ont été chargées par les cadres supérieurs de la Geography Division (le chef de la division et les chefs adjoints de la division) d'élaborer, à plein temps, un programme à long terme pour les travaux de soutien géographique du Bureau of the Census. Comme l'on s'y attendait, ce sont les problèmes sérieux, encore frais à notre mémoire, que nous avons connus lors de la livraison des produits et services cartographiques et géographiques pour le recensement de 1980 qui ont donné l'impulsion à ce travail.

Le petit groupe de planification disposait d'ébauches, produites peu de temps auparavant, de deux rapports d'évaluation des opérations du recensement de 1980. Ces rapports avaient été préparés par des groupes de travail inter-divisionnaires distincts composés de personnes qui avaient travaillé au recensement. Il découlait de ces deux rapports qu'un des principaux dénominateurs communs des diverses opérations était le besoin d'obtenir des produits de soutien géographiques et cartographiques cohérents et de les obtenir à temps.

Le premier résultat de notre effort de planification a été la production à temps d'un plan et d'un budget à présenter au comité de régie du Bureau of the Census en vue de la soumission du budget de l'exercice financier 1983. Faute de disposer de suffisamment de temps pour expliquer adéquatement ce programme ambitieux, nous n'avons pas réussi à obtenir l'approbation nécessaire. Des discussions ultérieures et une meilleure documentation à l'appui des avantages du programme proposé ainsi que de son effet sur le budget, de même qu'une participation plus active des fonctionnaires concernés du Department of Commerce et de l'Office of Management and Budget, ont entraîné l'approbation de notre programme à long terme et la promesse de niveaux de financement plus élevés à compter de l'exercice financier 1984 (soit du 1er octobre 1983). Ainsi, en moins de trois ans, nous avons réussi à obtenir l'autorisation de modifier de façon spectaculaire la façon dont la fonction de soutien géographique pourrait être réalisée pour les décennies à venir. C'est grâce à l'existence d'un plan d'action à long terme que nous avons pu atteindre cet objectif. Sans un effort de planification préalable, il y avait peu d'espoir de pouvoir lancer et concrétiser un programme géographique aussi ambitieux.

BIBLIOGRAPHIE

Broome, F.R., 1984. "TIGER Preliminary Design and Structure Overview, the Core of the Geographic Support System for 1990", réunion annuelle de l'Association of American Geographers de 1984, Washington, D.C.

Marx, R.W., 1984. "Developing an Integrated Cartographic/ Geographic Data Base for the United States Bureau of the Census", réunion de l'Austra Carto One, Perth, Western Australia.

FICHIER PRINCIPAL DE LA RÉGION

Un meilleur moyen de satisfaire les besoins des utilisateurs de données du recensement
Dans quelle mesure ce programme devrait-il être élargi?

JOEL YAN ET KAROLE KIDD

DIVISION DES SERVICES ET DU DÉVELOPPEMENT INFORMATIQUES
STATISTIQUE CANADA

JEAN-PIERRE PARKER

DIVISION DE LA GÉOGRAPHIE
STATISTIQUE CANADA

1. Introduction

Le présent document a pour objet de présenter le programme des fichiers principaux des régions (FPR) et son évolution depuis 1969. Il comporte une brève description du concept du FPR ainsi que les motifs sous-jacents à son élaboration. Suit un aperçu des techniques utilisées pour l'établissement et la maintenance d'un fichier principal de la région, puis un exposé de l'évolution du programme du recensement de 1971 à ceux de 1976, 1981 et 1986, c'est-à-dire des améliorations apportées jusqu'à ce jour. Cet exposé se termine par l'examen des principales questions concernant l'avenir des FPR et par certaines recommandations à ce chapitre.

2. Vue d'ensemble du programme

Depuis plus de 15 ans, la Division de la géographie de Statistique Canada travaille à l'élaboration et la mise à jour de fichiers numériques de réseaux de rues, connus sous le nom de fichiers principaux des régions. Ces fichiers servent de base géographique à un système plus vaste, le Système de stockage et d'extraction des données codées suivant une grille géographique (GRDSR), qui a pour fonction de permettre l'extraction des données du recensement selon des régions géographiques diverses.

2.1 Objet du FPR

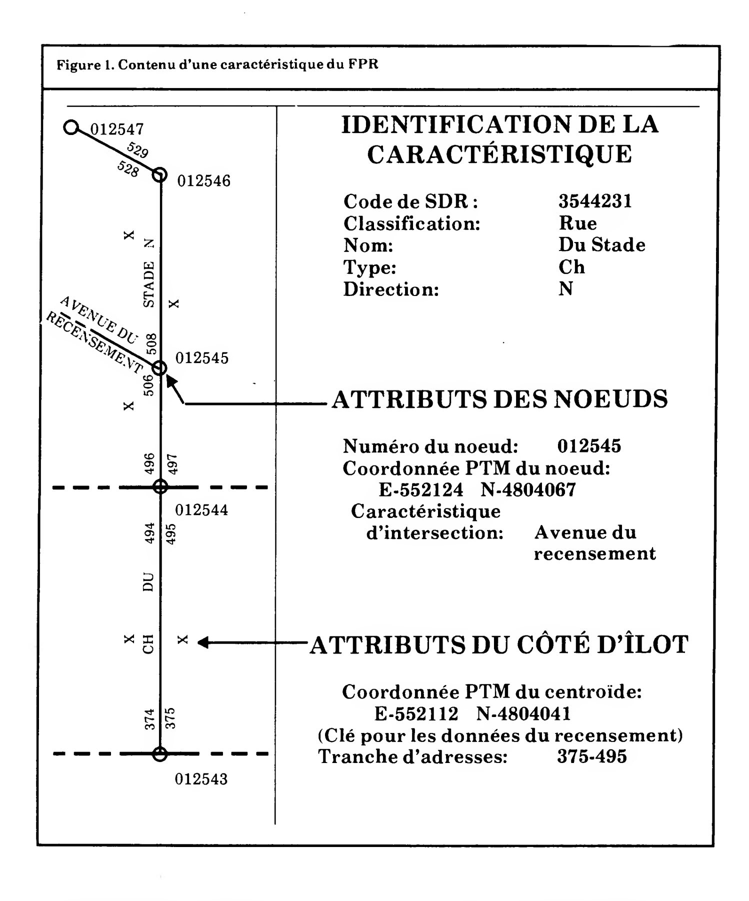
Le fichier principal de la région est un fichier géographique de base unique en son genre. Les données d'un FPR sont présentées par subdivision de recensement (SDR); un même FPR peut couvrir plus d'une SDR complète. Le FPR est un répertoire numérique de toutes les rues de ville et d'autres caractéristiques importantes (autoroutes, voies ferrées, cours d'eau, lignes hydro-électriques, etc.) qui servent à délimiter les régions particulières définies par les utilisateurs.

Chaque caractéristique du FPR est classée selon trois types de renseignements. Le premier correspond à la définition de la caractéristique, c'est-à-dire le code de la subdivision de recensement, le nom (par ex., rue Albert, Lac Dow) et la catégorie (par ex., rue, lac, île). Le deuxième est la représentation cartographique de chaque caractéristique sous forme d'une série de noeuds ou de points utilisés pour en délimiter, aussi précisément que possible, la configuration et l'emplacement à l'aide d'une simple ligne. Enfin, des tranches d'adresses (de voirie) sont établies pour chaque segment de la caractéristique (pour chaque côté d'îlot). Le côté d'îlot est un côté de rue situé entre deux intersections consécutives formées par la rencontre d'autres rues (voir la figure 1).

Au cours de l'établissement du fichier, le centroïde d'un côté d'îlot est automatiquement calculé et les coordonnées du centroïde sont utilisées pour représenter l'"étendue du côté d'îlot" (la tranche d'adresses d'un côté d'îlot donné). Ceci constitue la clef à l'aide de laquelle les données du recensement relatives aux ménages résidant sur le côté d'îlot peuvent être repérées géographiquement, ou "géocodées", selon les demandes spéciales des utilisateurs portant sur des secteurs particuliers, comme les zones scolaires, les districts de police, etc. Dans le cas des régions urbaines, l'utilisation des centroïdes de côté d'îlot a élargi d'environ 25 fois le champ géographique d'extraction des données par rapport à celui qu'il était possible d'obtenir à l'aide des centroïdes des secteurs de dénombrement (SD).

L'utilisation du centroïde de côté d'îlot permet également de fournir des données dont l'enchaînement chronologique est assez précis. Par contre, comme les secteurs de dénombrement sont susceptibles de varier de façon

Figure 1. Contenu d'une caractéristique du FPR



importante d'un recensement à un autre (jusqu'à 40%), la continuité chronologique à ce niveau géographique est pratiquement inexistante. D'autres méthodes visant à maintenir une plus grande stabilité, par exemple, en ayant recours à un îlot invariable en tant qu'unité géostatistique, sont présentement à l'étude (Parenteau, 1985).

2.2 Méthodes utilisées pour l'établissement ou la mise à jour du FPR

L'établissement d'un fichier principal de la région se fait à partir d'une série de plans de rues très précis provenant de diverses sources, principalement d'organismes municipaux (voir la figure 2). On utilise des plans récents, à grande échelle, sur lesquels sont indiquées les tranches d'adresses par côté d'îlot, de même qu'un répertoire des rues à jour.

Les plans manuscrits sont divisés en sections qui peuvent être introduites dans un convertisseur analogique/numérique, puis chaque intersection de deux caractéristiques, de même que le début, la fin ou un changement de direction important des caractéristiques sont marqués d'un noeud. Un numéro de série est ensuite attribué à chaque noeud. À l'étape suivante, les codes descriptifs de chaque caractéristique sont consignés sur une formule spécialement conçue à cette fin. Ces codes donnent le nom, le type, la direction, le nombre de noeuds et les numéros de voirie aux intersections de chaque caractéristique. Le plan manuscrit est alors prêt à être numérisé.

Le convertisseur numérique mesure l'emplacement des noeuds par rapport aux points de contrôle et attribue à chaque noeud une "coordonnée cartésienne" horizontale (est) (X) et verticale (nord) (Y) (voir la figure 3). Étant donné que le système de coordonnées utilisé pour le FPR est celui de la projection transversale de Mercator (PTM), les coordonnées cartésiennes doivent être converties en coordonnées PTM au cours d'un traitement sur ordinateur subséquent. Pendant ce traitement, les coordonnées PTM de chaque centroïde de côté d'îlot sont calculées à l'aide des coordonnées des noeuds se trouvant à chaque extrémité du côté d'îlot (voir la figure 4). Enfin, toutes les sections et tous les éléments sont regroupés et un fichier principal de la région est établi pour la subdivision de recensement.

Chaque fois qu'un fichier principal de la région est créé ou mis à jour, il fait l'objet d'un certain nombre de mesures visant à repérer et à corriger les erreurs. Une vérification approfondie par ordinateur est effectuée pour assurer que chaque noeud est attribué aux segments de rue pertinents, et l'inverse. Ceci permet de repérer la plupart des erreurs manuelles. Lorsque chaque section du fichier est terminée, elle est tracée à la même échelle que celle de la carte originale. Les deux cartes sont ensuite comparées afin de vérifier l'intégralité et l'exactitude du FPR. Tout tracé subséquent permet, après deux ou trois cycles de mise à jour, de produire un fichier principal de la région exempt d'erreurs. Lorsque les documents de base satisfont aux normes, le processus d'établissement d'un FPR pour une SDR d'environ 25,000 habitants requiert approximativement 40 jours-personnes répartis sur une période de 80 jours.

3. Évolution du FPR: 1969-1985

L'évolution du FPR de 1969 à nos jours est présentée en fonction de son développement permanent et du nombre croissant de produits et d'applications qui en résultent. Les progrès accomplis sont relatés dans une série de rapports portant sur l'étendue de la couverture du FPR au moment de chaque recensement.

3.1 Conception, élaboration et création initiale d'un FPR

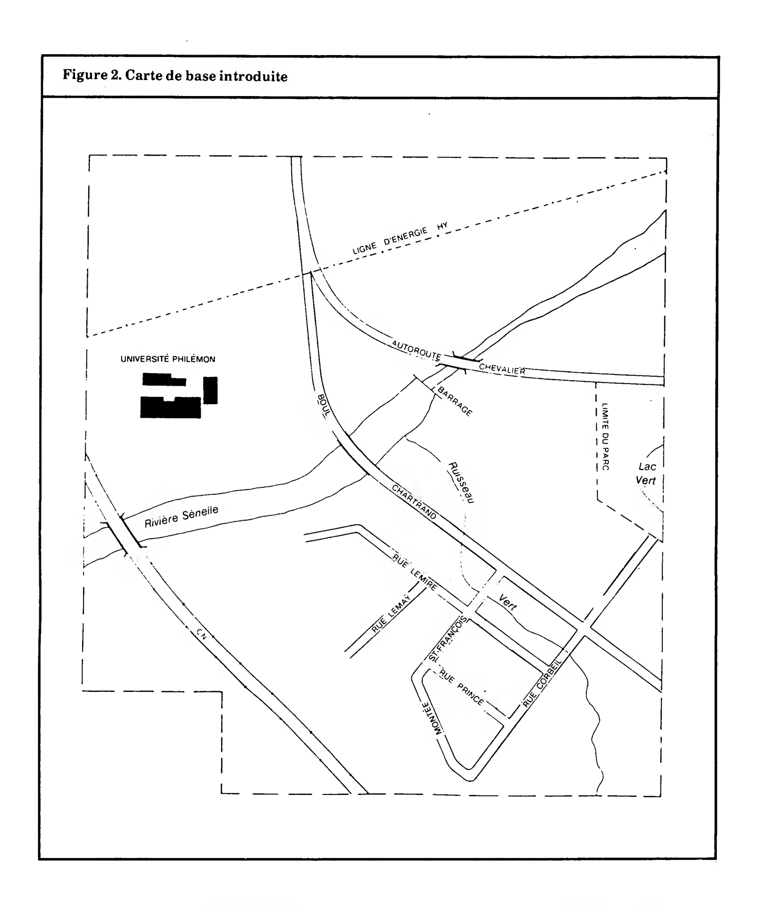
La conception et l'élaboration du logiciel du GRDSR et du FPR ont eu lieu au cours de la période 1969 à 1970 (Ion, 1969). Un prototype de FPR a été créé en 1969 pour la ville de London (Ont.) et a été éprouvé avec succès.

L'objectif à long terme était l'établissement de FPR permettant une couverture intégrale des subdivisions de recensement comprises dans l'ensemble des régions métropolitaines de recensement (RMR) et des agglomérations de recensement (AR) de plus de 50,000 habitants. Toutefois, au cours de l'étape initiale de la création des FPR, de 1970 à 1971, la priorité a été accordée à l'établissement de parties des 14 principales RMR et AR, selon les demandes des administrations municipales qui étaient prêtes à collaborer avec Statistique Canada. En conséquence, le système de géocodage et certains aspects du logiciel ont été confiés à plusieurs municipalités importantes qui les ont utilisés pour l'établissement de leur propre système de référence géographique (Richmond, 1974).

3.2 Recensement de 1971

Au cours du dépouillement des données du recensement de 1971, il a été possible

Figure 2. Carte de base introduite



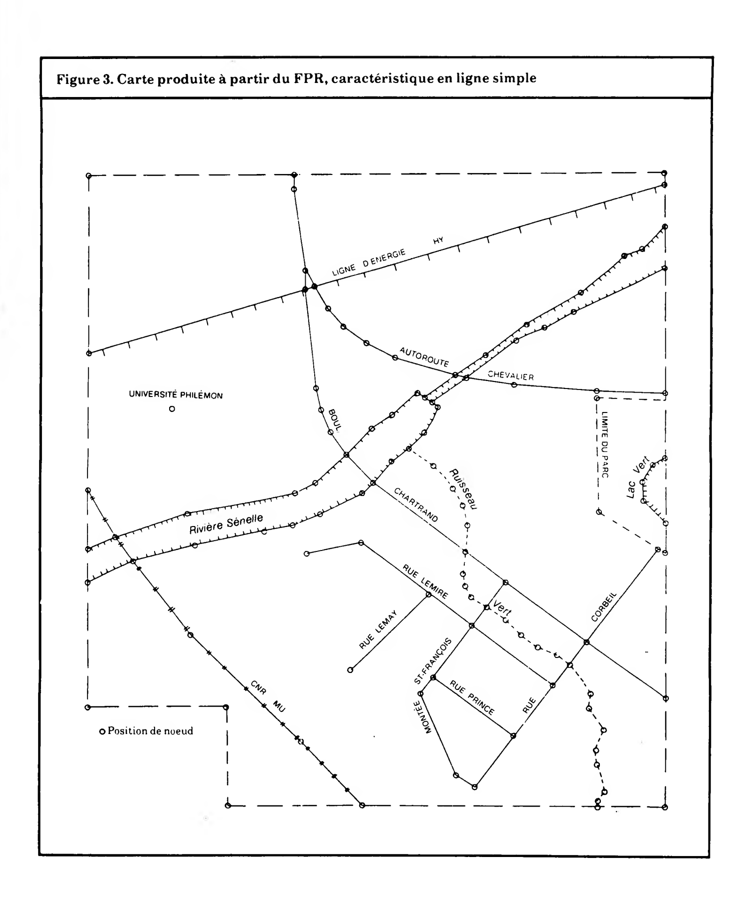


Figure 3. Carte produite à partir du FPR, caractéristique en ligne simple

Figure 4. Centroïde de côté d'îlot

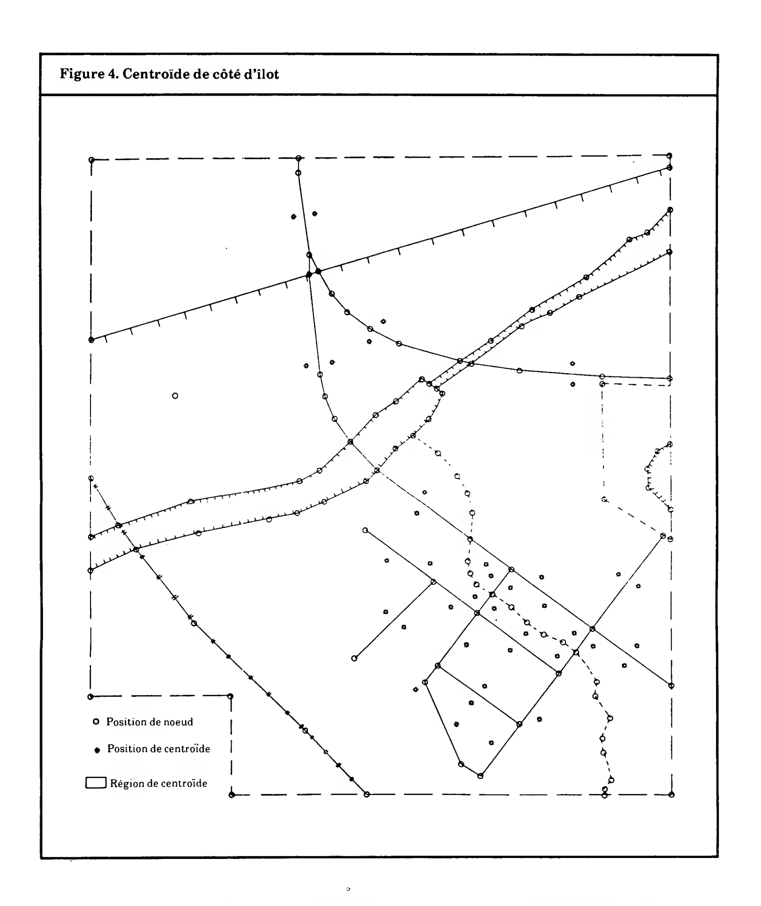


TABLEAU 1. Couverture des fichiers principaux des régions selon l'année censitaire

(Nombre de subdivisions de recensement, population couverte et pourcentage)

Province	Recensement de 1971			Recensement de 1976			Recensement de 1981		
	SDR	Pop.	%	SDR	Pop.	%	SDR	Pop.	%
T.-N.	-	-	-	1	86,576	15.5	1	83,770	14.7
Î.-P.-É.	-	-	-	-	-	-	-	-	-
N.-É.	1	122,035	15.5	2	183,223	22.1	2	176,871	20.9
N.-B.	-	-	-	1	85,956	12.7	2	135,264	19.4
Qc	12	1,677,611	27.8	58	3,277,026	52.6	59	3,233,410	50.2
Ont.	12	3,055,090	39.7	40	5,082,079	61.5	48	5,496,971	63.7
Man.	12	535,217	54.2	1	560,874	54.9	1	564,473	55.0
Sask.	2	265,918	28.7	2	283,343	30.8	2	316,823	32.7
Alb.	2	841,471	51.7	2	931,278	50.7	2	1,124,989	50.3
C.-B.	13	877,956	40.2	18	1,174,432	47.6	42	1,580,190	57.6
T. N.-O.	-	-	-	-	-	-	-	-	-
Yukon	-	-	-	-	-	-	-	-	-
Canada	54	7,375,298	34.2	125	11,664,787	50.7	159	12,712,761	54.5

d'attribuer un code géographique aux données sur les ménages résidant sur chaque côté d'îlot à partir des centroïdes de côté d'îlot enregistrés dans les FPR existants. Pour faire face à la demande croissante, des FPR ont été créés en 1973 pour les municipalités avoisinantes des villes de Vancouver et de Winnipeg, puis un code géographique a été attribué de façon rétroactive aux données du recensement de 1971 relatives à ces centres. Dans l'ensemble, le géocodage par côté d'îlot s'est étendu à 54 SDR au total et à environ 7.5 millions de personnes, soit 34% de la population canadienne visée par le recensement de 1971 (voir le tableau 1). Ces 54 SDR ont, de fait, été subdivisées en près de 200,000 microsecteurs, ce qui a rendu les opérations d'extraction et d'agrégation spéciales des données beaucoup plus flexibles qu'elles ne l'étaient (Renseignements par petits secteurs, 1972).

3.3 Élargissement important de la couverture du FPR

En prévision du recensement de 1976, un programme visant à élargir le géocodage par côté d'îlot a été élaboré de concert avec les municipalités. Grâce à la collaboration des municipalités, qui fournissaient les documents et assuraient les opérations de codage, alors que Statistique Canada s'occupait de la conversion numérique des cartes, du dépouillement et de l'assemblage des renseignements dans les FPR, il a été possible d'établir et de mettre à jour un plus grand nombre de FPR tout en maintenant les coûts à peu près au même niveau.

3.4 Recensement de 1976

Par suite de cet élargissement, l'étendue du géocodage par côté d'îlot a été accrue de 4.3 millions, soit 16% de la population, et a atteint 11.6 millions, ou 50.7% de l'ensemble de la population du Canada. Le nombre de subdivisions de recensement représentées s'est accru à 125 (voir le tableau 1), et ces SDR ont été subdivisées en approximativement 350,000 microsecteurs ou côtés d'îlot.

3.5 Nouveaux produits et applications du FPR

Une fois réalisé l'objectif initial, c'est-à-dire la production de données selon des régions délimitées par les utilisateurs, les responsables du programme sont passés à l'élaboration de nouveaux services fondés sur les données particulières du FPR. En même temps, la couverture du FPR a été élargie, en fonction des demandes spéciales, à des centres de moins de 50,000 habitants.

Répertoires des rues automatisés

Dans le passé, les répertoires des rues étaient établis manuellement, ce qui demandait beaucoup de temps et donnait lieu à de nombreuses erreurs. Ce travail a été automatisé pour répondre aux besoins du bureau du Directeur

général des élections et du personnel chargé des opérations du recensement sur le terrain. Dans le cas du bureau du Directeur général des élections, il était possible d'identifier rapidement la circonscription électorale fédérale où un électeur devait se rendre, simplement d'après l'adresse de l'électeur telle qu'elle figurait dans un dérivé du répertoire des rues. De même, pour ce qui est des opérations du recensement sur le terrain, il était très facile de déterminer à quel commissaire au recensement ou à quel recenseur il fallait confier le suivi lorsqu'un membre responsable du ménage se plaignait qu'il n'avait pas reçu de questionnaire à une adresse donnée.

Système prototype de production de cartes de dénombrement par ordinateur

La conception et l'élaboration d'un système prototype en vue de produire par ordinateur des plans de rue à double ligne à partir de la représentation à ligne simple contenue dans le FPR a permis de produire des cartes pour la collecte des données dans environ 34 SR (Yan, Bradley, 1982). Même si, au début, cet essai n'a permis de réaliser aucune économie, il a été prouvé qu'il était possible de produire de telles cartes et que ce processus pouvait devenir avantageux si certaines améliorations y étaient apportées.

Utilisation des FPR par d'autres organismes

D'autres organismes, comme les services de police de Stratford et le Engineering Department for road inventory de la ville de St. Catharines (Loewen, 1979), utilisent le FPR comme principal fichier cartographique.

3.6 Recensement de 1981

Pour le recensement de 1981, la couverture des FPR a été élargie à 159 SDR ce qui permet d'obtenir des renseignements géocodés par côté d'îlot sur environ 12.7 millions de personnes ou 54% de l'ensemble de la population du Canada (voir le tableau 1). Il est maintenant possible d'extraire des données régionales sur 18,000 SD, ou environ 500,000 micro-unités, et les ménages concernés.

3.7 Projets récents

Les récents développements permettent d'entrevoir un accroissement marqué des avantages que les applications des FPR présentent pour le recensement et les autres opérations de Statistique Canada. La coopération avec les organismes de cartographie fédéraux et provinciaux s'amplifie et les projets conjoints avec les administrations municipales et les autres organismes visant à élargir la couverture des FPR se multiplient. Voici une brève description de quelques-uns de ces projets.

Production de cartes de dénombrement pour le recensement de 1986

En 1983, un système de production de cartes de dénombrement à partir des FPR a été élaboré à l'aide du prototype décrit au paragraphe 3.5. Ainsi, 12,000 cartes de secteurs de recensement, représentant environ 6,000 SD, ont été produites par ordinateur pour le recensement de 1986, au coût de $60 par carte comparativement à $200 par carte dessinée manuellement.

En outre, ce système se prête à la production de cartes à double ligne comportant l'indication, pour chaque centroïde de côté d'îlot, du nombre de ménages et des codes postaux.

Fichier de conversion des codes postaux

Un logiciel a été conçu et mis sur pied avec succès en vue d'apparier les codes postaux (à 6 caractères) du fichier des codes postaux et les centroïdes de côté d'îlot correspondants, selon les tranches d'adresses du FPR. Ceci permet d'extraire des données du recensement selon des secteurs donnés à l'aide des codes postaux.

Un fichier de conversion des codes postaux (pour l'appariement des codes postaux et des régions géostatistiques normalisées de 1981) a également été élaboré, ce qui permet d'extraire des données sur certaines régions géostatistiques normalisées à partir des dossiers administratifs.

Fichier d'appariement des données par côté d'îlot

Un fichier spécial de données sur les ménages et la population par côté d'îlot a été produit pour l'enquête sur la population active en vue de la modification des unités d'échantillonnage.

Projets conjoints d'élargissement de la couverture

Depuis le recensement de 1981, des clients de l'extérieur ont témoigné un intérêt croissant pour les fichiers numériques de rues, ce qui a

amené l'entreprise de plusieurs projets conjoints visant l'élargissement de la couverture des FPR (Yan et Parker, 1985).

Un projet particulier a donné lieu à un important enrichissement des FPR dans le sud de l'Ontario, par suite duquel les services d'urgence peuvent maintenant utiliser ces derniers comme fichiers géographiques principaux et comme système cartographique de base pour l'acheminement par ordinateur des véhicules d'urgence. La plus grande partie du travail, qui a nécessité huit années-personnes, a été accomplie par les services de police.

Plusieurs projets ont été réalisés conjointement afin d'éviter le double emploi et de répartir le travail. Par exemple, la région métropolitaine de Toronto se propose de mettre à jour son fichier numérique du réseau de rues en fonction de la situation existant en 1985. Une fois révisé, le fichier servira à la mise à jour et(ou) au remplacement de la version de 1981 du fichier de Statistique Canada. D'autres projets semblables sont en cours de négociation avec les municipalités de Burnaby, Winnipeg (Courage et al., 1982) et Calgary (McNabb, 1982).

Un mémoire d'entente entre le ministère des Ressources naturelles de l'Ontario, le comté d'Oxford et la ville de Woodstock d'une part et la Division de la géographie de Statistique Canada d'autre part a été rédigé en vue de la création d'un FPR pour la ville de Woodstock.

Projet pilote avec la participation d'EMR

Un projet pilote a été entrepris avec la participation d'Énergie, Mines et Ressources (EMR). Le fichier numérique d'EMR contenant les valeurs des caractéristiques des coordonnées PTM a été chargé, puis les attributs nécessaires à la création du FPR y ont été ajoutés (Yan et al., 1985).

Extension de la formule de recouvrement des coûts

Des négociations sont en cours en vue de l'octroi de permis de vente de FPR en tant que formule de recouvrement des coûts d'élaboration et de mise à jour.

Dans ce but, le système de production de cartes de dénombrement a été élargi, ce qui a permis d'obtenir plus de $50,000 pour le travail de cartographie et d'établissement d'un répertoire des rues exécuté selon une formule de recouvrement des coûts pour le service ambulancier en Ontario.

De ce fait, on constate que le champ d'application s'est élargi dans une mesure telle que les mécanismes de production, les structures des fichiers et la capacité des systèmes existants sont sur le point d'atteindre leur limite, et qu'il serait approprié de marquer un temps d'arrêt et d'examiner les perspectives d'avenir des FPR.

4. Perspectives d'avenir des FPR

En 1969, lorsque le FPR a été conçu en tant que fondement du GRDSR, il était difficile de prévoir dans quelle mesure le FPR évoluerait comme produit unique et la diversité d'applications qu'il offrirait grâce à l'appariement avec d'autres fichiers. Le temps est venu, au milieu des années 1980, de réévaluer le FPR et de planifier son développement et son élargissement en vue du recensement de 1991 et d'autres utilisations. Il ne s'agit pas seulement de déterminer dans quelle mesure la couverture du FPR devrait être élargie, mais également d'examiner par quel moyen on pourrait le faire. Plusieurs scénarios et recommandations sont proposés ici.

4.1 Nécessité d'élargir le FPR

Les tendances récentes et l'expérience acquise au cours des applications particulières des FPR dans le contexte du recensement et d'autres projets font nettement ressortir la nécessité d'un élargissement constant de la couverture des FPR. Voyons plusieurs cas particuliers en guise d'exemples.

1. À l'heure actuelle, le FPR sert de base non seulement pour l'extraction des données du recensement, mais également pour le programme de cartographie du recensement, l'établissement de répertoires des rues et d'autres produits. Dans les cas où le système de production de cartes de dénombrement par ordinateur a été appliqué, il a été possible de faire des économies considérables (le coût d'une carte dessinée à la main, qui était de $200, a été réduit à $60 grâce à ce programme). De plus, la base servant à la production de cartes de dénombrement est compatible avec la base utilisée pour l'extraction des données du recensement, ce qui permet de réduire le nombre d'années-personnes requis pour le rapprochement manuel et la mise à jour de ces deux bases. La couverture actuelle du FPR limite toutefois les possibilités de faire de telles économies.

2. Les utilisateurs qui demandent des données selon des régions géographiques

particulières ont plus de facilité à les obtenir à partir de la couverture des FPR par côté d'îlot qu'ils ne le peuvent dans le cas des régions urbaines pour lesquelles il n'y a pas de FPR. À titre d'organisme fédéral, nous ne devons pas perdre de vue les répercussions politiques éventuelles liées à l'existence de ces deux niveaux de couverture (et, par conséquent, de service).

3. La fourniture de données du recensement selon le code postal promet d'être très rentable, bien que les possibilités offertes par l'appariement du code postal et du FPR n'aient pas encore été exploitées à fond. Le code postal pourrait fort bien être le genre d'indicateur géographique ayant la flexibilité requise pour adapter les données aux limites géographiques actuelles des villes, à condition que le fichier de conversion soit tenu à jour et qu'il y ait un FPR. Autrement dit, ce n'est que si la couverture des FPR était élargie qu'il serait intéressant d'utiliser les codes postaux (Puderer, 1985).

4. Il est recommandé que les FPR servent de base pour repérer les éléments d'adresse qui feraient partie d'un registre des adresses (Royce, 1985).

5. Des utilisateurs externes particuliers, comme les services de police et les services d'ambulance, utilisent les FPR pour l'acheminement par ordinateur de leurs véhicules et collaborent avec Statistique Canada à l'élargissement des FPR. De nouvelles demandes d'élargissement des fichiers sont présentées régulièrement par ces organismes.

4.2 Scénarios pour la gestion de l'élargissement de la couverture des FPR

Plusieurs points importants relatifs à l'élargissement de la couverture des FPR doivent être pris en considération, à savoir:

1. le champ d'application des FPR;

2. le rôle de Statistique Canada en ce qui concerne la production de FPR et la coordination de la production de FPR par d'autres organismes;

3. la mesure dans laquelle les FPR devraient être élargis, tant sur le plan logique que pratique;

4. le contenu des FPR par rapport aux besoins de Statistique Canada et à la structure des bases de données cartographiques produites par d'autres organismes;

5. le degré selon lequel le logiciel devrait être modifié.

Divers scénarios et recommandations sont proposés relativement à chacun de ces points. Il est à souhaiter que si on examine ces points maintenant, des solutions efficaces puissent être trouvées en ce qui concerne les possibilités futures des FPR pour le recensement de 1991 et au-delà.

Point n° 1: Champ d'application des FPR

Choix de scénarios

1.1 Utiliser les FPR pour les besoins du recensement seulement.

1.2 Utiliser les FPR pour toutes les activités de Statistique Canada.

1.3 Utiliser les FPR en fonction des besoins de Statistique Canada et des organismes de l'extérieur (principalement les services d'urgence, les organismes municipaux).

Recommandations:

1. Étendre le champ d'application des FPR dans le cadre du recensement, en utilisant le FPR comme principale base des régions géographiques urbaines. Les nouvelles applications du FPR pourraient comprendre le géocodage automatisé des données sur le lieu de travail, le programme d'îlots du recensement et la délimitation des terres agricoles pour le recensement de l'agriculture.

2. Promouvoir l'utilisation des FPR en tant que principal outil géographique pour d'autres opérations de Statistique Canada, par exemple, pour l'établissement d'un registre des adresses, pour le programme des données régionales, pour le programme d'utilisation des données administratives, pour l'appariement des codes postaux avec les FPR, et en faire la démonstration.

3. Élaborer des applications efficaces pour les utilisateurs de l'extérieur, selon une formule de recouvrement des coûts et à titre de "service public".

Point nº 2: Rôle de Statistique Canada dans l'élargissement des FPR

Choix de scénarios

2.1 Confier l'établissement des FPR au personnel du recensement seulement.

2.2 Confier l'établissement des FPR au personnel de Statistique Canada seulement.

2.3 Confier l'établissement des FPR au personnel du recensement et à celui de certains organismes de l'extérieur qui ont affaire à Statistique Canada.

2.4 Mettre sur pied un programme aux niveaux national et provincial pour recruter des associés qui seraient prêts à participer à l'élargissement des FPR.

Recommandations:

1. Les responsables du recensement devraient veiller à ce que l'élargissement des FPR réponde dans une proportion minimale à leurs propres besoins, à partir de leurs propres ressources.

2. Chaque fois que le FPR est destiné à servir des intérêts locaux, les responsables du recensement devraient, autant que possible, confier les travaux d'élargissement du fichier à des organismes ayant une meilleure connaissance de la région, à condition que ces derniers aient la compétence technique nécessaire. Il est recommandé d'avoir, autant que possible, recours aux ressources locales.

3. Statistique Canada devrait encourager ce genre de participation en prévoyant:

   a) des trousses de documents à jour à l'intention des responsables et des préposés à la création de FPR locaux;

   b) des outils et le logiciel de base nécessaires à l'application des FPR en vue d'en accroître l'utilité pour d'autres organismes.

4. Les fonctions principales de Statistique Canada devraient être les suivantes:

   a) établir les normes minimales en ce qui concerne le contenu des fichiers et la compatibilité de ces derniers;

   b) organiser le travail et les communications;

   c) servir d'organe centralisateur pour l'échange d'information et de logiciel.

Point nº 3: Limite de la couverture des FPR

Choix de scénarios

3.1 Ne pas modifier la couverture des FPR.

3.2 Élargir la couverture des FPR à un niveau géographique constant, selon les ressources existantes, par exemple, en fonction des RMR ou AR entières, d'une population établie des SDR, ou des tournées des facteurs.

3.3 Élargir la couverture des FPR en fonction de la demande et des ressources, uniquement à l'intention du marché extérieur.

3.4 Élargir la couverture des FPR selon une formule basée sur la demande et l'utilité.

3.5 Élargir progressivement la couverture des FPR à l'échelle nationale, dans le cadre de projets conjoints avec les principaux associés (comme dans le cas du TIGER).

Recommandations:

1. Le groupe de la géographie devrait consulter les responsables du recensement et d'autres programmes de Statistique Canada pour déterminer la nécessité d'élargir les FPR en vue d'y inclure d'autres régions du pays: priorité 1, priorité 2, priorité 3, etc.

2. Une norme d'utilité devrait être établie pour l'élargissement des FPR en fonction des avantages que peut en tirer Statistique Canada et des coûts que cela représenterait pour le Bureau dans le cas où des organismes locaux ou autres seraient prêts à partager ces coûts ou à constituer un fonds à cette fin.

3. Une liste des principaux organismes qui sont sur le point, ou qui se proposent, d'élaborer d'importants programmes de cartographie numérique devrait être dressée. Des relations devraient ensuite être établies avec les organismes dont le

contenu et la précision des programmes sont acceptables à Statistique Canada. Des programmes conjoints applicables à l'échelle nationale devraient alors être élaborés, puis la jonction entre ces programmes devrait être établie et leur faisabilité éprouvée aussi rapidement que possible. L'élargissement des FPR en vue du recensement de 1991 devrait être planifié. La participation des cadres supérieurs devrait être assurée.

Point nº 4: Mesure dans laquelle le contenu des FPR devrait être modifié

Choix de scénarios

4.1 Immobiliser le contenu des FPR à son état actuel.

4.2 Apporter des améliorations mineures au contenu des FPR au fur et à mesure des besoins.

4.3 Être prêt à élargir le contenu des FPR en fonction des besoins des principaux utilisateurs et associés.

Recommandations:

1. Maintenir les concepts de base du fichier des réseaux et prévoir une interface entre ce fichier et le format actuel des FPR, quelles que soient les améliorations apportées à leur contenu.

2. Poursuivre le processus d'enrichissement progressif du contenu des FPR en vue de permettre l'amélioration des éléments actuels, par exemple, les principales routes et les principaux immeubles.

3. En prévision du recensement de 1991, entreprendre une révision approfondie de l'infrastructure géographique en vue d'améliorer la liaison entre les divers niveaux (réseaux, limites, codes postaux) de l'infrastructure et rentabiliser la mise à jour et le traitement. Un examen approfondi du système TIGER s'impose également.

4. Lorsque des associés auront été trouvés pour le programme d'élargissement, s'il y a lieu, il conviendrait de tenir compte, au cours des discussions portant sur les améliorations du contenu des FPR, des services que ces autres organismes seraient prêts à fournir en contrepartie. Le fichier des réseaux devrait toutefois être maintenu en tant que base principale, avec des couches distinctes pour les attributs, codes postaux, limites, îlots, etc.

Point nº 5: Utilisation de nouvelles techniques pour le programme de maintenance des FPR

Choix de scénarios

5.1 Pour ce qui est du système de maintenance, adopter une attitude réactive, c'est-à-dire conserver le système existant, sauf en ce qui concerne la correction des erreurs.

5.2 En ce qui a trait au développement du système, adopter une attitude proactive.

Recommandations:

1. Les méthodes et(ou) les techniques devraient être rentabilisées tout en tirant profit des derniers progrès techniques afin d'être en mesure d'en faire plus avec les ressources existantes.

5. Sommaire des recommandations

Dans la section précédente, nous avons fait un exposé assez détaillé des divers points en rapport avec le développement constant des FPR. Voici une récapitulation des principales recommandations.

1. Maintenir le soutien du programme des FPR à son niveau actuel.

   Compte tenu des succès obtenus grâce à l'élargissement des FPR sur le plan de la géographie du recensement et d'autres projets, assurer des crédits permanents de façon que le programme de base des FPR puisse au moins être maintenu.

2. Poursuivre le développement des applications des FPR.

3. Élargir la couverture des FPR à l'aide des ressources permanentes de Statistique Canada et des projets conjoints avec les principaux associés.

4. Passer en revue le contenu des FPR et l'améliorer en fonction des besoins des principaux utilisateurs et associés afin de faciliter l'utilisation des fichiers.

5. Élaborer un plan à long terme pour l'utilisation des FPR.

Compte tenu de l'évolution des avantages et des profits qu'offre l'utilisation des FPR, un plan à long terme devrait être élaboré pour le développement de ces fichiers. Un tel plan devrait principalement inclure:

a) l'analyse des nouvelles applications des FPR par les responsables du recensement, de Statistique Canada et des organismes de l'extérieur;

b) une enquête exhaustive auprès d'autres organismes s'occupant de cartographie numérique; et

c) l'élaboration de lignes directrices pour faciliter et soutenir les "projets conjoints".

6. Augmenter le budget des services de recherche et de développement des FPR.

   Les services de recherche et de développement devraient:

   a) améliorer l'infrastructure;

   b) élargir la série de produits;

   c) rentabiliser les systèmes et les méthodes; et

   d) apporter leur aide aux projets conjoints.

7. Élaborer un programme de promotion et de mise en marché.

BIBLIOGRAPHIE

Boisvenue, A., et Parenteau, R., 1981. "The Geocoding System in Canada and its Area Master File", **exposés présentés à la conférence annuelle de l'URISA**, p. 226 - 232.

Courage, W.G., et Bennett, P., 1982. "A Practical Information System Developed With Limited Funds", **exposés présentés à la conférence annuelle de l'URISA**, p. 159 - 172.

Division de la géographie, Statistique Canada, 1983. "Champ d'observation du géocodage au niveau du côté d'îlot, recensements de 1971, 1976 et de 1981", document de travail nº 6.

Ion, R.J., 1969. "The Geographic Basis of the DBS Geocoding System for Urban Areas: An Overview", note de service analytique et technique nº 3, Bureau fédéral de la statistique, Ottawa.

McNabb, G.H., 1982. "Spatial Data, Geographic Referencing and Computer-assisted Mapping at the City of Calgary", **exposés présentés à la conférence annuelle de l'URISA**, p. 233 - 241.

Parenteau, Robert F., 1986. "Un programme d'îlots - Oui ou non?" **Compte rendu de la conférence internationale sur la planification du recensement de 1991**, Statistique Canada.

Puderer, Henry, A., 1982. "Census Geography Staff's Planned Postal Code Master File - Area Master File Linkage", document rédigé dans le cadre de l'atelier sur le système canadien de géocodage lors des séances tenues par l'Association canadienne des géographes à l'occasion de la conférence des Sociétés savantes de 1982, Université d'Ottawa.

Puderer, Henry, A., 1984. "Postal Codes and the 1981 Canadian Census of Housing and Population", **actes de la conférence de l'Urban and Regional Information Systems Association**, qui s'est tenue à Seattle (Washington) du 12 au 15 août 1984, version révisée: 84/8/30.

Puderer, Henry, A., 1985. "Statistics Canada's Postal Code Products and Services", **actes de la conférence de l'Urban and Regional Information Systems Association**, qui s'est tenue à Ottawa (Ontario) du 28 juillet au 1er août 1985.

Richmond, D.E., 1974. "The Joy of George - A Municipal Guide to Geocoding", Management Systems Development Department, City of Calgary, 16 pages.

Royce, Don, 1986. "Applications d'un registre des adresses au recensement canadien", **Compte rendu de la conférence internationale sur la planification du recensement de 1991**, Statistique Canada.

Statistique Canada, juin 1972. **GRDSR: Renseignements par petits secteurs**, manuel d'introduction, Ottawa.

Yan, J.Z., et Bradley, D.R., 1983. "Computer-assisted Cartography for Census Collection: Canadian Achievements and Challenges", **Les actes du sixième Symposium international sur la cartographie automatisée**, Washington, mars 1985, p. 584 - 599.

Yan, Joel Z., et Parker, Jean-Pierre, 1985. "A Framework for Coordinating the Development and Application of Street Network Files for Canada", **exposés présentés à la conférence annuelle de l'URISA**, p. 132 - 143.

UN PROGRAMME D'ÎLOTS

Oui ou non?

ROBERT PARENTEAU

DIVISION DE LA GÉOGRAPHIE
STATISTIQUE CANADA

Introduction

Le développement de la structure géographique sous-tendant la tenue d'un recensement de la population et du logement est une opération très complexe. Pour les besoins de la collecte des informations de tous les ménages canadiens, l'ensemble du territoire est subdivisé en petites unités, le secteur de dénombrement (SD). Ce dernier représente l'unité géographique dénombrée par un recenseur et **il change à chaque recensement**.

L'instabilité du secteur de dénombrement (SD) d'un recensement à l'autre constitue l'élément principal de cette présentation. Ce changement constant est une source d'inefficacité, pour les raisons suivantes:

1. difficulté de conduire des études longitudinales car les bandes sommaires sont préparées pour les SD, de même que la méthodologie d'évaluation "contre-vérification des dossiers" qui consiste à retracer les ménages au recensement précédent;

2. obligation de recodifier complètement les filières géographiques (particulièrement le Fichier géographique principal du recensement), lesquelles sont basées sur le SD;

3. réduction de la flexibilité dans la description des territoires définis par les utilisateurs des données statistiques;

4. fort impact sur d'autres activités menées à Statistique Canada (ex., enquête sur la population active);

5. imposition de contraintes opérationnelles allant à l'encontre du désir commun de décentraliser dans les bureaux régionaux la délimitation des secteurs de dénombrement;

6. incapacité de contrôler les changements (ces derniers résultant de modifications aux critères établis pour la délimitation des SD par les opérations régionales de révisions aux unités géostatistiques, et de la demande de certains utilisateurs pour changer les limites des SD); et

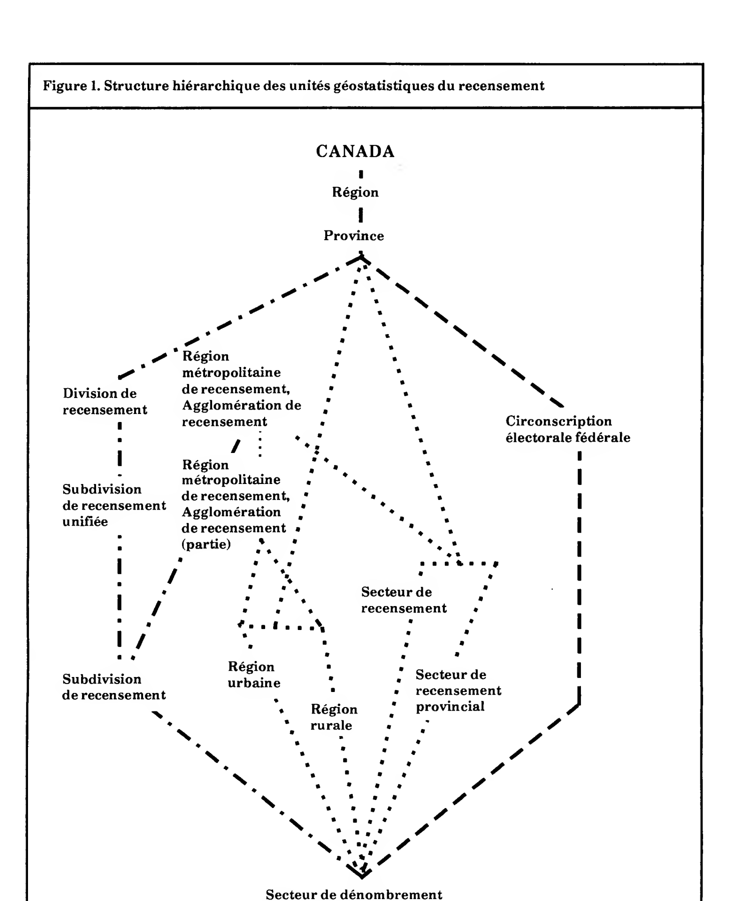
7. retenue imposée à la délimitation optimale de certaines unités géostatistiques (tels les SR/SRP, les régions urbaines) car les opérations de collecte demandent de maximiser les limites des SD, de ce fait apportant des changements opérationnels de limites aux unités géostatistiques (tableau 1).

Tableau 1. Nombre de SD* selon la méthodologie de 1986

Méthodologie	>		<		Moyenne	Total
Retour par la poste (critère: 375 log.)	2,511	12%	18,851	88%	284.70	21,362
Reprise – Urbain (critère: 300 log.)	658	12%	4,762	88%	160.06	5,420
Reprise – Rural (critère: 175 log.)	3,068	20%	12,185	80%	109.25	15,253
Total	6,237		35,798			42,035

*85-09-30

Figure 1. Structure hiérarchique des unités géostatistiques du recensement



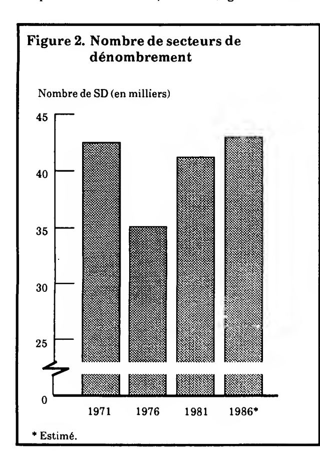
I. Expérience passée

Afin de mieux saisir l'importance du secteur de dénombrement, nous pouvons regarder la figure 1 qui démontre où se situe le SD au sein des unités géostatistiques ainsi que le rôle vital qu'il joue dans la compilation de données.

Pour la délimitation des SD, on procède par circonscription électorale fédérale (CÉF) qui est la division territoriale élisant un député à la Chambre des communes. Les CÉF sont modifiées après chaque recensement décennal pour tenir compte des changements de distribution de la population; leurs limites peuvent déborder les divisions de recensement, les subdivisions de recensement, mais jamais les provinces. Le recensement de 1991 devrait être basé sur une nouvelle ordonnance de représentation.

Comme le SD doit respecter les limites des diverses unités géostatistiques, lesquelles peuvent subir plusieurs changements durant la période intercensitaire, ce dernier est très instable. Lors des trois derniers recensements, le nombre de SD a fortement fluctué: 42,533 en 1971, 35,154 en 1976, 41,197 en 1981; pour 1986, on prévoit environ 43,000 SD (figure 2). Cette variation est en grande partie le résultat de changements des critères de délimitation des SD. Le réseau des SD ne jouit donc pas de la stabilité suffisante pour assurer l'obtention, à partir d'un recensement donné, des informations concernant un territoire quelconque, dans les limites qu'il avait aux recensements précédents, quand ces limites ne sont plus les mêmes. Les SD d'un recensement donné ne sont pas, en général, superposables à ceux du ou des recensements précédents. Leur délimitation n'observe pas non plus les limites des unités géostatistiques telles qu'elles étaient au(aux) recensement(s) précédent(s), si ces limites ont changé. L'origine de l'instabilité du SD est son usage multiple: (1) unité administrative et opérationnelle lors de la collecte des données sur le terrain, (2) unité statistique pour la diffusion des données. Chaque fonction suppose un certain nombre d'exigences particulières, fondamentales, parfois divergentes.

Figure 2. Nombre de secteurs de dénombrement



Afin de solutionner ce problème d'instabilité, il a été décidé, pour le recensement de 1986, que les limites des secteurs de dénombrement ruraux aussi bien que celles des régions urbaines seraient les mêmes que celles du recensement de 1981 dans tous les cas où il serait possible. Suivant cette dernière politique, on visait la stabilité des limites des SD pour le recensement de 1986, comparativement à celles de 1981. Les travaux de préparation pour le recensement de 1986 nous ont fait perdre plusieurs de nos illusions. Le secteur de dénombrement, de par ses multiples fonctions, n'est pas une unité stable car seulement 62% des SD de 1986 ont des limites comparables aux limites de 1981.

II. Alternatives au secteur de dénombrement

Le secteur de dénombrement, tel qu'il existe, ne répond pas aux exigences de la fonction de diffusion de données que sont la stabilité et la flexibilité.

Le retrait de la fonction de diffusion contribuerait à faciliter la création et la délimitation des secteurs de dénombrement. Il s'agit donc de trouver une solution de substitution. Toute solution possible dépend de facteurs et de contingences divers, tels que: temps, personnel et budget disponibles; possibilité ou non d'utiliser des éléments existant déjà; d'utiliser des méthodes ou de l'équipement déjà en production; ou encore la possibilité de reprendre les unités d'un recensement précédent (ex., SD 1986) comme point de départ d'un nouveau système rendant possible la comparaison de données à partir de ce recensement.

Diverses solutions peuvent être considérées. En voici quelques-unes présentées sommairement.

(i) Emploi d'un quadrillage régulier: ce dernier serait générateur d'aires géographiques permanentes capables de servir d'éléments de reconstitution des entités de diffusion dans l'élaboration des données statistiques.

(ii) Emploi des SD des recensements précédents comme unité de base: les SD des recensements précédents seraient géocodés et deviendraient la base de diffusion des données pour les recensements à venir. Ces derniers seraient ainsi figés dans un réseau d'unités permanentes, superposables à chaque recensement aux nouveaux SD et servant de base de comparaison chronologique. Ces unités de base permettraient de reconstituer les entités de diffusion.

(iii) Emploi du code postal: la Grande-Bretagne étudie présentement la possibilité d'utiliser le code postal comme unité de collecte de l'information. Comme le système postal couvre tout le pays et que les codes postaux sont rattachés aux unités de diffusion, ce dernier pourrait devenir l'unité opérationnelle de la collecte.

(iv) Délimitation d'îlots: le Bureau du recensement américain a établi comme objectif pour le recensement de 1990 de découper le territoire américain en îlots. Ce dernier est une extension de programmes mis en place au cours des deux derniers recensements (1970 et 1980). Le recensement canadien utilise déjà le concept d'îlot, même si une définition claire n'a pas encore été développée.

La première solution permettrait la délimitation d'unités stables, flexibles et permanentes. Par contre la mise en place d'un tel système serait sûrement coûteuse. Cette nouvelle technique apporterait une brisure considérable dans l'évolution de la structure géographique canadienne. Je ne crois pas qu'on soit prêt à une telle évolution. De plus, l'intégration d'une grille régulière avec les limites irrégulières des îlots formant les unités géostatistiques causerait énormément de difficultés.

La deuxième solution serait beaucoup moins révolutionnaire. Le principe de construction par élément prédéterminé existe déjà avec le secteur de dénombrement. L'unité définie serait stable, flexible et applicable au Canada entier.

La troisième solution, qui consiste dans l'utilisation du code postal comme unité de collecte, pourrait être retenue là où existent des fichiers principaux des régions (c.-à-d. pour les grands centres urbains canadiens). Dans ces centres, l'appariement "code postal et côté d'îlot" a été développé. Par contre, cette méthode ne peut s'appliquer aux petits centres urbains sans des investissements énormes pour créer d'autres FPR. Pour le milieu rural, cette méthodologie est impensable car un code postal rural couvre un large territoire. Cette solution ne doit pas être rejetée complètement; une application par étape pourrait offrir des gains potentiels.

La dernière solution, retenue par le Bureau du recensement américain, consiste en la répartition du territoire en îlots. Dans les pages suivantes, nous allons référer à cette option, en comparant avec l'expérience américaine.

La solution ou la combinaison de solutions retenue doit préserver autant que possible la continuité historique, évitant ainsi une cassure nette dans la chronologie statistique. Une telle rupture serait susceptible de susciter des difficultés de raccordement dans l'espace et dans le temps, tel que satisfaire les besoins justifiés des utilisateurs de raccorder leurs séries statistiques basées sur les SD (des bandes sommaires, par SD, ont été annoncées et vendues depuis le recensement de 1971).

La délimitation d'îlots pourrait être la solution la plus efficace. Une première ébauche de conceptualisation et de délimitation d'îlots urbains a été élaborée lors des recensements de 1971, 1976, 1981 et 1986.

III. Expérience des recensements de 1971 à 1986

La définition d'îlots n'est pas vraiment une idée nouvelle. La Division de la géographie avait considéré cette question en 1971 et 1976 où l'îlot de recensement était présenté comme un outil défini par Statistique Canada pour rendre disponible l'information du recensement sur le plus petit territoire possible, tout en respectant les restrictions prescrites par la Loi sur la statistique. Au recensement de 1971, l'îlot était défini comme la plus petite superficie entourée de rues. Un îlot était généralement rectangulaire mais il pouvait être aussi de forme irrégulière et avoir des limites d'un autre genre, par exemple, un cours d'eau ou une limite politique. Au recensement de 1976, la définition d'îlot se précise: un îlot de recensement est la plus petite unité fermée qui peut être délimitée à l'intérieur d'un secteur de

recensement ou d'un secteur de recensement provincial, et peut varier en terme de population, de limites, de forme ou de dimension (figure 3).

Une évaluation des îlots, pour les recensements de 1971 et 1976, a déjà été faite. Les points suivants sont un résumé des constatations les plus importantes de cette évaluation.

(i) L'utilisation première du numéro d'îlot était reliée aux opérations de collecte et de traitement de données. Les agents recenseurs devaient procéder îlot par îlot; les numéros d'îlot étaient inscrits dans le Registre des visites, facilitant l'accomplissement d'activités géographiques suivant immédiatement le jour du recensement.

(ii) L'utilisation du numéro d'îlot pour la diffusion était considérée comme un objectif secondaire.

Suite aux expériences de 1971 et de 1976, il avait été recommandé pour le recensement de 1981 que seulement les subdivisions de recensement faisant partie du programme de secteur de recensement soient subdivisées en îlots. Cette politique a été suivie, mais l'îlot n'a jamais été défini comme une unité géostatistique. Le peu d'importance accordée à l'îlot est facile à constater; le Dictionnaire du recensement de 1981 ne renferme aucune définition de ce dernier. Certaines données du recensement (population et nombre de ménages) sont disponibles par îlot mais aucune publicité ou mise en marché n'en a été faite. Les restrictions budgétaires et les coupures affectant la Division de la géographie à la fin des années 1970 ont retardé le développement de tout nouveau concept géostatistique.

À l'été de 1982, il a été décidé que pour plus de 1,200 secteurs de recensement compris dans les fichiers principaux des régions (FPR), la délimitation des îlots (et des SD) serait faite automatiquement à partir des traits incorporés aux FPR. Le FPR est un document en langage lisible à la machine, lequel décrit le tracé des rues ainsi que d'autres traits physiques à l'aide d'un système de coordonnées.

Pour les cartes de collecte du recensement de 1986 produites automatiquement, les îlots sont donc déterminés à partir de tous les traits linéaires des FPR, sauf les limites de propriété, les gares de triage et les voies ferrées d'accommodement. Les numéros d'îlots sont placés à l'intérieur de ces derniers, à l'endroit défini par les centroïdes. Les numéros sont assignés à chaque îlot en ordre croissant (ex., 1, 2, 3). Si l'îlot est trop petit, le numéro n'est pas tracé automatiquement. Les numéros d'îlots ne sont donc pas parfaitement en séquence; des numéros d'îlots sont également assignés aux lacs, rivières et îles. Les numéros sont assignés d'ouest en est, du sud au nord.

La production automatique des cartes de secteurs de recensement et la délimitation/numérotation des îlots permettent de réduire de beaucoup les coûts de préparation des cartes de secteurs de recensement. Ainsi, le coût estimé pour produire automatiquement les cartes de SR est de $60 par carte, comparativement à $200 pour les cartes produites manuellement. À moins d'imprévus, la production automatique de cartes de collecte devrait être plus importante encore pour le recensement de 1991. Tous les secteurs de recensement (environ 40,000) devraient être produits automatiquement, à partir des FPR. Vu les efforts soutenus pour améliorer le contenu des FPR et pour accroître son champ d'application, la délimitation automatique des îlots devrait englober une forte proportion de la population canadienne en 1991 (environ 65%).

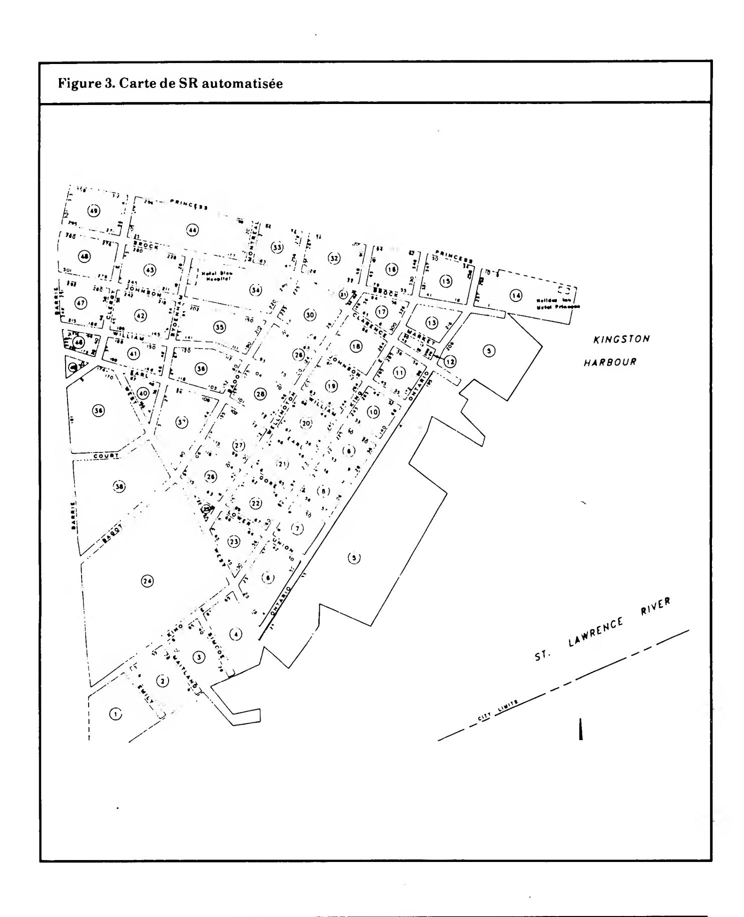
IV. Stratégie pour les prochains recensements

Des investissements importants en temps et en argent ont été consacrés au développement d'îlots lors des derniers recensements mais ces investissements n'ont pas toujours été productifs faute d'une politique bien définie.

La question qui se doit d'être posée à l'intérieur de cette conférence est la suivante: Est-ce que Statistique Canada devrait financer d'autres travaux afin de déterminer le bien-fondé d'un programme d'îlots? Si on se base sur l'expérience du recensement américain, des îlots ont été définis et délimités depuis 1970 comme complément aux districts de dénombrement. Pour le recensement de 1990, la politique adoptée par le Bureau du recensement américain est que le programme d'îlots s'étendra à l'ensemble de la nation américaine. Les îlots et les districts de dénombrement seront définis ou bien comme unité de diffusion ou comme unité administrative, lesquelles seront mutuellement exclusives. Toutes les régions pour lesquelles les données seront disponibles par îlot ou par groupe d'îlots auront des districts de dénombrement comme unité administrative et vice-versa.

Cette dernière notion est très importante. Elle est la réponse à une des contraintes rattachées aux secteurs de dénombrement, l'instabilité, laquelle dépend du double rôle qu'on veut faire jouer au SD. Définissant une unité de diffusion et une unité de collecte/administrative indépendante

Figure 3. Carte de SR automatisée



l'une de l'autre, il devient plus facile de contrôler les changements affectant chaque type d'unités.

Les experts du recensement américain, à partir de l'expérience du recensement de 1980, reconnaissent que le contrôle des activités de collecte au niveau de l'îlot permet une meilleure supervision de l'ensemble de la tâche et une amélioration de la couverture nationale. Pour les utilisateurs, la disponibilité des données, pour l'ensemble du pays, au niveau de l'îlot offre la flexibilité d'agréger ces données pour leur territoire d'intérêt.

Cette notion de flexibilité d'agrégation des données par les utilisateurs est une question primordiale: Est-ce que les utilisateurs désirent obtenir les données du recensement par îlot ou préfèrent-ils maintenir le statu quo? Pour l'utilisateur conditionné à obtenir les informations statistiques par secteur de dénombrement, l'introduction d'une nouvelle unité de diffusion pourrait être une source importante de problèmes et de confusion.

Afin de mieux connaître la réaction des utilisateurs, il y aura lieu de les consulter. Une première consultation pourrait être faite à partir des bureaux régionaux, lesquels maintiennent des contacts réguliers avec les utilisateurs des statistiques. Si on veut que l'information statistique soit plus utilisée, l'utilisateur doit sentir que cette information peut l'aider dans ses décisions de tous les jours.

Le Bureau du recensement américain, afin de parer aux réactions négatives des utilisateurs locaux, a encouragé les organisations locales à soumettre des propositions de délimitation de districts de dénombrement et d'îlots. Le programme des secteurs de recensement, pour le recensement du Canada, fonctionne sur le même principe. Ces mêmes contacts pourraient devenir les coordonnateurs pour un programme d'îlots à l'intérieur des grands centres urbains.

Cette dernière question est vraiment la plus importante: un programme d'îlots sera élaboré seulement s'il répond à des besoins exprimés par les utilisateurs de données du recensement. Si les utilisateurs préfèrent le statu quo, une unité spéciale pourrait être définie uniquement pour la collecte des données. Elle ne deviendrait pas une unité de diffusion.

Si la réaction des utilisateurs est que le programme d'îlots apporterait beaucoup à la population canadienne, Statistique Canada devrait élaborer une politique nationale. Le recensement de 1991 verrait la réalisation d'un projet-pilote où toutes les interrogations possibles seraient examinées: une définition d'îlot, les limites choisies, la numérotation, la concordance avec les SD des recensements précédents, la diffusion de données par îlot ou par groupe d'îlots, la confidentialité des données statistiques. De plus, Statistique Canada devra profiter de l'expérience américaine. Des discussions intensives devront être menées avec nos confrères du recensement américain.

D'autres perspectives doivent également être envisagées. Quels sont les avantages pour l'ensemble du Bureau de la mise sur pied d'un tel programme? **L'enquête sur la population active** pourrait bénéficier énormément d'une numérotation permanente des îlots, lesquels constituent l'élément de base de l'échantillonnage. L'échantillon retenu présentement est basé sur les îlots (et numéros d'îlots) du recensement de 1981. Afin de mettre à jour le nombre de logements par îlot, pour refléter les résultats du recensement de 1986, l'EPA doit relier les îlots de 1986 aux îlots de 1981. Une numérotation permanente faciliterait l'appariement et diminuerait les coûts d'une telle opération.

La préparation des cartes de collecte, par la Division de la géographie, pourrait devenir une opération moins coûteuse suite au développement d'un tel programme. Les coûts de la collecte pourraient-ils être réduits? Est-ce que les bureaux régionaux pourraient définir eux-mêmes les secteurs de dénombrement par le groupement d'îlots? Toutes ces questions devront être examinées avant de pouvoir proposer un tel programme.

SYSTÈME INFORMATIQUE DE SOUTIEN DE LA GÉOGRAPHIE DU RECENSEMENT

GORDON DEECKER, RON CUNNINGHAM ET KAROLE KIDD

DIVISION DES SERVICES ET DU DÉVELOPPEMENT INFORMATIQUES
STATISTIQUE CANADA

1. Introduction

La structure de base de tout recensement moderne est le système géographique mis sur pied pour la circonstance. Ce fait est souligné par Gordon (1975) selon qui: (traduction libre) "un recensement doit s'appuyer sur un cadre spatial en fonction duquel les données sont recueillies, totalisées et présentées". Marx (1985) reprend la même idée en termes généraux: (traduction libre) "Le succès d'un recensement dépend non seulement de l'efficacité de la collecte des données mais également de la manière dont nous relions ces dernières aux régions géographiques". Tomasi (1985) relève trois outils géographiques principaux servant à étayer les données du recensement, à savoir:

- les cartes
- les fichiers de référence des adresses
- les fichiers de référence géographiques

Les **cartes** relatives à la géographie du recensement comprennent les cartes de référence qui illustrent le cadre géostatistique, les cartes de dénombrement qui sont des guides pour la distribution et la collecte des questionnaires du recensement et les cartes thématiques qui illustrent des thèmes particuliers pour un ensemble donné d'unités géostatistiques.

Les **fichiers de référence des adresses** contiennent des représentations numériques des réseaux de rues et des caractéristiques associées ainsi que les tranches d'adresses, pour chaque côté de rue, comprises entre des caractéristiques consécutives. Ces fichiers sont appelés fichiers GBF/DIME aux États-Unis, pour le recensement de 1980, et fichiers principaux des régions (FPR) à Statistique Canada.

Les **fichiers de référence géographiques** présentent, sous forme numérique, les liens entre les divers éléments de la hiérarchie des unités géostatistiques. Le Master Reference File (MRF) aux États-Unis et le Fichier géographique principal du recensement (FGPR) au Canada en sont des exemples.

Les éléments géographiques et cartographiques du recensement sont appuyés par des systèmes informatiques qui doivent être utilisés pour traiter les ensembles de données spatiales considérables et complexes que nécessite tout recensement moderne.

Dans le présent document, nous étudierons les systèmes spatiaux qui ont été utilisés pour le recensement de 1971 et les pressions dont ils ont été la cible. La nature changeante du processus de collecte des données et des exigences des utilisateurs de statistiques entraîne la nécessité d'apporter des modifications importantes à ces systèmes. Nous tenterons de montrer comment il est possible de profiter de ces exigences pour améliorer les systèmes.

2. Systèmes de la géographie du recensement à Statistique Canada

L'élaboration des systèmes géographiques destinés à étayer le recensement du Canada a commencé à la fin des années 1960 quand l'application de la technologie informatique aux données spatiales en était à ses débuts. Les écrits de Ion (1969), de Podehl (1971) ainsi que le manuel du GRDSR publié par Statistique Canada (1972) contiennent des renseignements sur les systèmes qui sont encore utilisés aujourd'hui. Au début des années 1970, les fichiers de données du FGPR ont été élaborés pour compléter la couverture des données urbaines et pour fournir une couverture uniforme pour tout le pays dans le cas de certaines unités géostatistiques.

Le Système de stockage et d'extraction des données codées suivant une grille géographique (GRDSR) a été élaboré pour le recensement de 1971 comme moyen d'extraire des données du recensement pour toute région définie par l'utilisateur, que cette région corresponde ou non à une ou plusieurs unités géostatistiques normalisées. Le processus d'extraction est basé sur une série de centroïdes (précisés par leurs coordonnées), définis au niveau du côté d'îlot dans le cas des régions qui figurent dans les fichiers principaux des régions et au niveau du SD partout ailleurs au Canada. Ces centroïdes sont liés aux renseignements appropriés dans la base de données du recensement. Quand un utilisateur demande des renseignements pour un ensemble donné de régions, il les décrit sur une carte. Ces zones

spécifiées sont converties sous forme numérique et le système GRDSR identifie les centroïdes. Les statistiques sont alors totalisées dans un rapport.

Un fichier principal de la région répertorie les rues, les tranches d'adresses, les côtés d'îlot et les coordonnées de centroïdes dans les villes de 50,000 habitants ou plus. On y trouve aussi d'autres caractéristiques telles que les voies ferrées, les cours d'eau et les limites des villes. Les fichiers principaux des régions ont été conçus pour permettre l'extraction de données dans le cas de régions non normalisées et ils sont maintenant utilisés pour produire les cartes de dénombrement pour certains centres urbains principaux pour 1986. L'historique, le contenu et les applications des fichiers principaux des régions sont décrits dans le document de Parker, Yan et Kidd (1985).

Les buts principaux du FGPR sont les suivants: (1) fournir une base de données informatisée centrale de renseignements géographiques pour le recensement et (2) assurer la qualité des données géographiques en permettant une présentation et un regroupement adéquats des données du recensement. Les renseignements sont codés dans trois fichiers: le fichier des données géographiques regroupées (EAMF) qui contient les coordonnées cartésiennes de chaque secteur de dénombrement (SD), à titre de données de référence, et une série de codes identifiant les unités géostatistiques dans lesquelles le SD se trouve, le fichier des noms de lieux (ANAM) qui donne la Classification géographique type et le nom officiel de chaque région et qui contient des renseignements supplémentaires tels que la superficie et les chiffres de population provenant du recensement antérieur, et le fichier des limites (ABND), qui définit les polygones limites de certaines unités géostatistiques.

Depuis 1976, les systèmes de la géographie du recensement ont fonctionné avec des "modifications minimales", et la recherche sur de nouveaux systèmes a eu une portée limitée. Les raffinements et les améliorations effectués ont eu une importance mineure par rapport aux travaux de recherche et d'élaboration initiaux et ils ont été apportés pour répondre à des exigences particulières des utilisateurs ou à des pressions de l'extérieur. Les exigences ont porté principalement sur des formats diversifiés pour la présentation des données géographiques et des programmes machine distincts ont été élaborés à la suite des diverses demandes.

En résumé, à partir d'un seul système GRDSR, nous en sommes arrivés à un groupe éclectique de systèmes élaborés pour résoudre des problèmes particuliers ou pour répondre à des besoins d'amélioration particuliers. Collectivement, ces programmes constituent un "système de renseignements spatiaux" dont la pierre angulaire sont les fichiers principaux des régions et le Fichier géographique principal du recensement.

3. Points de pression

Quand des systèmes sont laissés en place pendant 15 ans, ils font l'objet de nombreuses pressions. Ces dernières sont causées en partie par les progrès technologiques et en partie par des changements au niveau de l'application qui en est faite. Nous avons défini les catégories suivantes pour traiter des pressions exercées sur les systèmes de géographie du recensement:

- Nouveaux produits
- Nouveaux clients
- Nouveaux systèmes
- Nouvelles idées
- Nouveaux projets conjoints
- Nouvelles données

Nous allons examiner ces points l'un après l'autre en donnant un exemple concret dans chaque cas.

3.1 Nouveaux produits

La demande de nouveaux types de produits géographiques dépasse ce que certains des systèmes existants peuvent offrir et entraîne la double nécessité d'élargir les bases de données et de modifier les structures des données. Voici certains des produits demandés récemment:

- cartes de secteurs de tri d'acheminement/cartes pour les codes postaux
- cartes pour les actions en cas d'urgence
- modélisation de données pour le tourisme
- cartes de données pour des côtés d'îlots
- cartes avec symboles multiples.

Le Trillium Data Group a obtenu, du ministère de la Santé de l'Ontario, un contrat pour produire un système d'acheminement des ambulances pour Halton-Peel Central Ambulance Dispatch Services (Services centraux de répartition des ambulances des comtés de Halton et de Peel). Trillium a passé un sous-contrat avec Statistique Canada en vue d'élargir le fichier principal de la région dans les comtés de Halton et de Peel et de produire des cartes de référence et un répertoire des rues pour cette région. Pour répondre à ces besoins, il a cependant fallu définir un nombre accru de types de rues dans la structure du fichier principal de la région.

3.2 Nouveaux clients

Un des éléments du programme de recouvrement des coûts du recensement de 1986 a été l'accent mis sur la création de sous-produits qui peuvent être élaborés de façon rentable et vendus à des clients de l'extérieur. Au cours des 18 derniers mois, un certain nombre de clients du secteur privé ont communiqué avec Statistique Canada afin d'obtenir des produits imprimés et des fichiers de données tirés de la base de données géographiques du recensement. Citons parmi ces entreprises:

- Gandalf Data Ltd.
- Trillium Data Group
- John Deere and Company
- Canadian Tire Corporation

Statistique Canada a récemment fourni à Gandalf Data Ltd. un répertoire des rues produit à partir du fichier principal de la région pour un service d'acheminement des taxis dans la région métropolitaine de Toronto. Cela permet d'introduire dans le système l'adresse d'un client en vue d'obtenir le nom des rues transversales des deux côtés, c'est-à-dire d'identifier l'adresse du client au côté d'îlot près.

En 1984-1985, les fichiers numériques des limites de municipalité pour l'ensemble du Canada (données utilisées à des fins d'assurance de la qualité dans le Fichier géographique principal du recensement (FGPR) - voir la section 2) ont été fusionnés et convertis sous une forme utilisable pour la production de cartes thématiques. Le fichier ne contient actuellement que les limites officielles qui ne correspondent pas au littoral (par ex., les limites peuvent se prolonger dans l'eau et il arrive souvent que plusieurs îles sont incluses dans un même polygone). L'addition du littoral à ce fichier augmenterait la diversité de ses applications, mais elle augmenterait aussi de façon substantielle le volume de données dans les fichiers de données du FGPR.

3.3 Nouveaux systèmes

De nouveaux systèmes fournissent de nouvelles occasions d'automatiser les traitements et de changer les structures de données existantes, ce qui permet de surmonter les problèmes causés par un excès de données et la rigidité des systèmes. Depuis l'élaboration du GRDSR pour le recensement de 1971, des améliorations importantes ont été apportées aux systèmes de renseignements spatiaux. Voici quelques systèmes typiques:

- CARIS
- ARC/INFO
- GBF/DIME
- TIGER
- ARIES III

Le système ARC/INFO, élaboré par l'Environmental Systems Research Institute, est un système de renseignements géographiques et de gestion de base de données d'avant-garde. Il permet d'effectuer, sous diverses formes, l'introduction et la manipulation des données; il permet aussi les consultations par l'utilisateur et la sortie, sous diverses formes, de données géographiques, ce qui n'est pas possible avec les systèmes utilisés actuellement par Statistique Canada. Ce système permet la subdivision et la manipulation de données géographiques au moyen de "tuiles" rectangulaires et il permettrait d'ajouter une structure topologique au FPR.

Clayton (1980) décrit comment les systèmes d'analyse d'images sont utilisés, pour traiter les données obtenues par satellite, afin de créer des systèmes de classification et d'utilisation des terres, pour détecter les changements dans l'utilisation des terres aux abords des régions métropolitaines et pour obtenir des estimations de la population dans les régions qui connaissent une croissance marquée. Le système ARIES III est actuellement utilisé par Statistique Canada pour effectuer l'analyse des données agricoles.

Les nouveaux systèmes sont vendus comme des outils qui permettent d'augmenter la productivité tout en diminuant les coûts. Si l'on se contente d'introduire sporadiquement ces systèmes dans l'ensemble des systèmes utilisés, pour remplacer certains systèmes qui s'y trouvent déjà, on aboutira à une augmentation du nombre des divers systèmes informatiques dont il faut assurer le fonctionnement.

3.4 Nouvelles idées

Pendant toute la durée d'un système, de nouvelles idées jaillissent en rapport avec les produits et les applications. Ces idées peuvent, à leur tour, engendrer des pressions visant à modifier les structures de base des données du système afin d'améliorer la puissance de traitement et la flexibilité de celui-ci. Ainsi, Statistique Canada envisage notamment:

- l'intégration du code postal à la base de données;

- l'addition, au FPR, de concepts topologiques pour traiter les îlots;
- l'intégration des fichiers de données du FPR et du FGPR;
- la cartographie des SD;
- l'introduction, la mémorisation et la sortie de données cartographiques selon des unités de grille normalisées;
- le couplage avec le projet de l'Atlas électronique.

Actuellement, les rues et autres caractéristiques linéaires du FPR sont représentées sous une forme qui ne contient aucun renseignement à propos des régions adjacentes (par ex., îlots, unités géostatistiques). D'autres fichiers géographiques représentent les unités géostatistiques sous forme de polygones, ce qui ne permet pas l'identification des régions environnantes. L'addition d'un élément topologique, comme dans le système TIGER, permettrait de décrire une limite ou un segment de rue dans une "représentation à 1 cellule" tout en incluant les renseignements connexes à propos de son lien avec d'autres noeuds (représentation à 0 cellule) et les régions (représentation à 2 cellules) qui se trouvent de chaque côté de cette limite ou de ce segment de rue. Ceci permettrait d'effectuer facilement la création et le groupement des régions, ainsi que d'autres opérations géographiques, cartographiques et analytiques.

Bien que l'addition d'éléments topologiques augmenterait de façon marquée la valeur des fichiers de données du FPR, cela entraînerait la modification d'un des éléments de base de la construction de ce système de renseignements spatiaux qui étaye la géographie du recensement.

La Direction des levés et de la cartographie d'Énergie, Mines et Ressources Canada (EMR) est en train d'élaborer un "Atlas électronique" pour avoir plus de flexibilité lors de la manipulation, de l'analyse et de l'utilisation créative des renseignements contenus dans l'Atlas national du Canada. Les fichiers de l'Atlas électronique ne sont pas simplement un recueil des cartes existantes; on y trouve aussi des données sur les positions et les attributs de façon que la création de nouvelles cartes soit possible par la manipulation, l'analyse et l'affichage interactifs des données. EMR est intéressé à relier les données et les limites géographiques du recensement avec les fichiers de données de l'Atlas. L'effort afin d'élaborer de nouveaux formats de présentation pour les fichiers du recensement conçus principalement à des fins d'assurance de la qualité peut entraîner des problèmes de programmation et d'autres problèmes d'ordre méthodologique.

3.5 Nouvelles données

Pour la majorité des accords de coparticipation conclus jusqu'ici il a fallu accroître les FPR. Un accord de ce genre a récemment été conclu avec la ville de Woodstock et le gouvernement de l'Ontario afin de créer un FPR pour Woodstock. Bradley (1985), dans une note de service interne, signale que de telles réussites pourraient se reproduire pour des "centaines" d'autres localités. Cela aurait pour effet d'ajouter une quantité considérable de nouvelles données à un système déjà surchargé.

L'augmentation du volume de données géographiques disponibles exerce une pression sur les besoins de traitement, pression qui dépasse parfois les limites du système ou qui réduit l'efficacité de certains programmes de traitement et de traçage conçus pour des bases de données beaucoup moins importantes. Ce type de situation a généralement été réglé au moyen de solutions "impromptues", ce qui laisse beaucoup à désirer du point de vue efficacité.

Les points suivants illustrent certaines des pressions qui entraînent la production de nouvelles données:

- la quantité de données du FPR a plus que triplé depuis 1971;
- les accords avec Woodstock et d'autres organismes pourraient la faire doubler à nouveau;
- les municipalités de 20,000 à 50,000 habitants désirent avoir leur propre FPR;
- l'augmentation de la précision implique l'utilisation d'une quantité accrue de données.

Le tableau 1 montre la croissance du nombre d'articles dans la base de données du FPR pour chacun des recensements depuis 1971. Il est prévu que d'ici 1991, le nombre d'articles dans la base de données augmentera de 500% par rapport à 1971.

Le Conseil national de recherches a démontré que le FPR constitue une base de données efficace pour les services de police. Le Geographic Resource Allocation Software System (GRASS), décrit par Arnold (1985), a été bien reçu. Actuellement, un groupe d'étude en

Tableau 1. Fichiers principaux de région		
Année	Nombre d'articles	% d'augmentation
1971	293,000	
1976	590,000	101
1981	721,000	22
1986	1,000,000	39
1991	2,000,000	100

recommande l'utilisation dans 120 localités en Ontario. Un bon nombre de ces localités comptent moins de 50,000 habitants. Des pressions seront donc exercées sur Statistique Canada afin d'accroître la base de données du FPR pour qu'elle comprenne ces régions.

En 1981, la ville de St. Catharines a passé un contrat avec Statistique Canada afin d'améliorer la qualité du FPR pour que son service des Travaux publics puisse l'utiliser. Dans l'ensemble, les améliorations ont été effectuées pour augmenter la précision de l'emplacement des noeuds, ce qui permet d'utiliser des calculs portant sur la longueur des rues dans les études d'urbanisme. Le résultat net de ces travaux a été d'augmenter le volume du FPR pour St. Catharines de 300%.

3.6 Nouveaux projets conjoints

Les projets entrepris conjointement avec d'autres organismes entraînent l'obligation de modifier les spécifications relatives à la structure, aux systèmes et au traitement. Les projets conjoints, en voie de réalisation ou à l'étude, englobent des organismes tels que:

- EMR
- Cambridge
- Woodstock
- IST (CRAR)
- Région métropolitaine de Toronto
- Burnaby
- Winnipeg
- Calgary

Statistique Canada, par l'intermédiaire de la Sous-division de la géocartographie, participe au projet d'Atlas électronique parrainé par EMR, en fournissant les services de programmeurs pour faire d'autres travaux d'élaboration afin que l'on puisse atteindre les objectifs communs et produire des interfaces simplifiées pour l'échange des données.

Depuis 1979, la ville de Winnipeg a utilisé le FPR de Statistique Canada comme source d'adresses de voirie pour l'acheminement des véhicules d'urgence (voitures de pompiers, de police et ambulances). De plus, le FPR est utilisé pour attribuer des coordonnées à chaque parcelle de terrain, ce qui permet d'effectuer des analyses spatiales à des fins de planification. La ville de Winnipeg effectue maintenant toutes les mises à jour des fichiers de la région et Statistique Canada utilise ces mises à jour numériques. Les travaux de traçage sont actuellement effectués par Statistique Canada.

Les échanges du type de ceux qui ont été conclus avec Winnipeg permettent à Statistique Canada d'obtenir des données de qualité améliorée. Ils imposent aussi des restrictions sur l'importance des modifications qui peuvent être apportées aux systèmes sans contrevenir aux accords conclus. Toute modification des formats utilisés pour les données et pour l'échange de données nécessiterait des changements aux systèmes utilisés par Statistique Canada et par la ville de Winnipeg.

3.7 Autres types d'incidence

D'autres éléments de nature politique ou économique ont un effet sur l'avenir du système informatique de soutien de la géographie du recensement. La décentralisation de la charge de travail en constitue un exemple. Les bureaux régionaux voudraient avoir un rôle accru dans la définition des unités géostatistiques, y compris les secteurs de dénombrement et les secteurs de recensement. Cette tendance à la régionalisation de la charge de travail peut entraîner l'obligation de modifier les systèmes.

Le coût du logiciel augmente en pourcentage du coût total d'un système. Ceci donne une importance accrue au partage du logiciel et à l'élaboration de systèmes à usage plus général.

Les progrès technologiques récents joueront un rôle majeur dans la façon dont les systèmes géographiques futurs seront structurés. La

baisse du coût du matériel nous permet d'augmenter l'automatisation de certains traitements en utilisant du matériel sophistiqué, par ex., des lecteurs. La disponibilité accrue des micro-ordinateurs entraînera une plus grande décentralisation, c.-à-d. plus d'analyses de statistiques par chaque utilisateur, ce qui provoquera une demande de logiciels géographiques spécialisés pour micro-ordinateur et de données statistiques et de données de position sur disquettes.

3.8 Récapitulation

Beaucoup des points de pression dont nous avons parlé sont interreliés. Les nouveaux clients veulent de nouveaux produits qui exigent des données additionnelles. Les nouvelles idées et les nouveaux produits entraînent des modifications dans les structures de données. Les projets conjoints entraînent de nouvelles idées et de nouvelles données.

Manifestement, le rôle des fichiers tenus pour la géographie du recensement change. Les fichiers de données du FPR sont utilisés dans tout le pays comme base pour des systèmes de renseignements spatiaux sur les municipalités. Les FGPR utilisés auparavant à des fins d'assurance de la qualité sont maintenant considérés comme une base pour la cartographie thématique de données du recensement. Cependant, les systèmes qui permettent de traiter ces fichiers n'ont pas été adaptés à l'évolution de la demande.

La rapidité et la puissance accrues du matériel informatique actuel font qu'il est possible d'envisager des applications qui, lors de la conception du GRDSR, n'étaient pas réalisables au point de vue économique. C'est le cas, par exemple, de la cartographie au niveau du SD. Si cette idée est acceptée, elle entraînera l'intégration des données conservées séparément dans les bases de données du FPR et du FGPR, ce qui nécessitera la modification de la structure des données.

4. Études connexes

Une analyse des publications depuis 1983 indique qu'il y a un point commun entre les problèmes de Statistique Canada et ceux qui touchent d'autres systèmes.

Peuquet (1984) constate que: (traduction libre) "l'étendue des données, en croissance rapide ... ainsi que le besoin de les utiliser ensemble ... ont révélé deux problèmes sérieux ... le fait que la gamme d'applications soit très peu étendue ... et l'inefficacité de la mémorisation et de la rapidité de traitement pour les volumes de données actuels et prévus".

Les problèmes relatifs aux données jouent un rôle important. White (1983) écrit que: (traduction libre) "le partage des données entre des personnes qui font des travaux de cartographie assistée par ordinateur pose tellement de problèmes que l'on n'y a généralement pas recours". Cette idée est partagée par Tomlinson (1984) dans son discours-programme lors du International Symposium on Spatial Data Handling, à Zurich, Suisse, quand il dit que: (traduction libre) "un souci récent est la possibilité d'échanger des données entre les bases; ... par exemple, pour transférer des données entre deux des principales bases géographiques au Canada (CANSIM et CGIS), il est encore plus rentable de tracer et de numériser les données à nouveau plutôt que d'effectuer un transfert numérique".

Les problèmes relatifs au traitement des données géographiques découlent de trois causes principales que Peuquet (1983) résume de la façon suivante: (traduction libre)

- "les limites ont tendance à être très compliquées et irrégulières";
- "les données numériques ont tendance à être incomplètes, imprécises et sujettes à erreur";
- "les liens spatiaux ont tendance à être flous ou limités aux applications spécifiques, et le nombre de liens spatiaux possibles est considérable".

Il est évident que le volume de la base de données est aussi un problème fondamental. Tomlinson (1984) écrit que: (traduction libre) "peu d'entre nous ont une idée de la dimension que prendront, ou auront à prendre, les bases de données". Actuellement, l'USGS est en train de numériser quelque 1,800 cartes pour le système TIGER. Au Canada, environ 1,000 cartes à l'échelle 1:250,000 sont en cours de numérisation. Cependant, les projets futurs sont beaucoup plus ambitieux. L'USGS prévoit numériser 55,000 cartes, opération qui pourrait produire 1.5x10**15 pixels de renseignements.

Les problèmes étant communs, nous devrions aussi étudier les recommandations proposées pour les élaborations futures afin de voir si nous pouvons adopter une base commune. Collins et al. (1983) suggèrent trois domaines où des améliorations peuvent être apportées:

- conversion à des systèmes de base de données relationnelles;
- amélioration du contenu par l'addition de données topologiques;
- amélioration au niveau des interfaces humaines avec les systèmes.

Nuttall et Korenstein (1985) exposent brièvement le projet de Geographical Referencing and Information System (GRIS) pour la région métropolitaine de Toronto, lancé en 1984. Ce document nous fournit des renseignements valables parce que l'organisme chargé de l'urbanisme pour la région métropolitaine de Toronto utilise des données provenant à la fois de CANSIM et du FPR. Le but visé est de relier cinq bases de données, contenant au total 964 milliards d'octets de données, au moyen d'un cadre spatial commun.

Le système GRIS est conçu pour remplir six fonctions de base, y compris:
(traduction libre)

- "permettre que des données tirées de sources avec des identificateurs géographiques différents soient classées ensemble dans une seule source avec un identificateur géographique commun";
- "faciliter l'analyse des données soumises à des contraintes relatives aux corridors géographiques ou à des aires de recrutement";
- "produire des relevés cartographiques à toute échelle".

Ces fonctions sont aussi des exigences de base des systèmes spatiaux de Statistique Canada. Compte tenu du fait que les données du FPR constituent un élément essentiel du système utilisé à Toronto, il est prévu que des consultations bilatérales sur les problèmes relevés et leurs solutions fourniraient des lignes directrices valables pour les futurs systèmes spatiaux de Statistique Canada.

5. Réactions possibles

Nous pouvons réagir de trois façons aux pressions exercées sur les systèmes actuels qui étayent la géographie du recensement:

- bloquer les systèmes existants;
- améliorer les systèmes existants;
- adopter de nouveaux systèmes.

5.1 Blocage des systèmes existants

Le fait de bloquer les systèmes existants permettrait d'effectuer le minimum de travail requis afin de pouvoir continuer à exploiter les systèmes actuels. Selon nous, cette décision serait inacceptable pour de nombreuses raisons, entre autres:

- nous éprouvons des difficultés avec les systèmes actuels;
- cela empêcherait de prendre de nombreuses autres initiatives dans le domaine de la géographie;
- on ne peut faire abstraction des pressions du marché;
- si la quantité de données augmente, les coûts de traitement vont devenir excessifs;
- c'est une question de crédibilité.

Dans une étude des systèmes spatiaux, Boisvenue et al. (1983) ont relevé deux problèmes importants. Tout d'abord, un bon nombre des programmes créés au cours des 12 dernières années ont été écrits dans des langages de programmation différents avec écriture permanente des options relatives au stockage des données et aux possibilités de traitement. De ce fait, il faudra écrire à nouveau ces parties du système, ne serait-ce que pour garder le système en exploitation. Deuxièmement, le rapport fait remarquer que notre possibilité de produire des données à partir des recensements actuels et antérieurs, selon un cadre géographique stable, ne progresse pas à la vitesse souhaitée par les clients.

5.2 Amélioration des systèmes existants

Une amélioration des systèmes existants permettrait d'augmenter l'automatisation des traitements individuels, car on pourrait alors remplacer les procédures manuelles par des procédures automatisées. Pour ce faire, il faut soit augmenter les possibilités des systèmes existants, soit élaborer (ou acquérir) un système additionnel.

Il convient de tenir compte de certains facteurs-clés dans les discussions portant sur l'amélioration des systèmes existants comme réponse valide aux pressions exercées sur les systèmes; en effet,

- certains systèmes ne peuvent être améliorés, il faudra les remplacer;
- l'amélioration de chaque système ne permet pas d'obtenir facilement une solution globale;
- pour les systèmes utilisés actuellement, il est impossible d'ajouter des liens spatiaux en se contentant d'apporter des améliorations aux systèmes.

Boisvenue et al. (1983) ont écrit que: (traduction libre) "la plupart des fichiers de données ont été élaborés pour une application particulière et l'on s'est peu préoccupé de leurs possibilités d'application plus universelle". Bien qu'il s'agisse d'une critique assez sévère pour un système qui a été utilisé pendant 15 ans, cela souligne la difficulté d'apporter des améliorations aux systèmes afin d'accroître la gamme d'applications possibles.

5.3 Adoption de nouveaux systèmes

L'adoption de nouveaux systèmes permettrait d'apporter des modifications aux systèmes ou d'utiliser de nouveaux systèmes de façon à dépasser les exigences minimales prévues actuellement d'après nos estimations des besoins en 1991.

Un changement est toujours un risque calculé qui suppose en général un double effort. Il faut conserver les systèmes actuels comme mesure de précaution pendant que les nouveaux systèmes sont élaborés et mis à l'essai.

Pour ce qui est des systèmes informatiques de soutien de la géographie du recensement, il existe des moyens de minimiser les risques courus. Le principal d'entre eux est ce que nous pouvons appeler la méthode "saute-mouton". Le tableau 2 donne une liste des divers progrès relatifs aux systèmes informatiques qui se sont produits aux États-Unis et au Canada au cours des 20 dernières années.

Tableau 2. "Méthode saute-mouton"

Recensement	États-Unis	Canada
1960	FOSDIC	
1966		Répertoires des rues
1970	Guide de codage des adresses DIME	
1971		FPR Extraction de secteurs non normalisés
1980	GBF/DIME	
1981		Production de cartes de dénombrement par ordinateur
1990	TIGER	
1991		

Comme nous pouvons le voir dans ce tableau, chaque pays a tiré des renseignements précieux du travail effectué par son voisin. Les idées et les procédures ont été adaptées aux besoins différents de chaque bureau du recensement. Cependant, la recherche importante qui a produit ces idées et ces procédures a été partagée plutôt que d'être reprise dans les deux pays. Par son empressement à envoyer des délégués parler du système TIGER et en faire la démonstration, le U.S. Bureau of the Census a démontré son désir de partager les travaux de recherche qui ont précédé l'élaboration de ce système. Cette connaissance, dûment adaptée au contexte canadien, réduira de façon appréciable le risque associé à l'autorisation de modifications importantes des systèmes informatiques pour le recensement de 1991.

6. Résumé

La géographie, la cartographie et les systèmes informatiques connexes sont des éléments de base du programme du recensement. Les systèmes actuels sont utilisés depuis 1971 et répondent à peine aux besoins. White (1983) cite l'aphorisme: (traduction libre) "Ce n'est pas ce que vous ne savez pas qui fait mal, c'est ce que vous savez ne pas être adéquat". Pour ce qui est des travaux du recensement, cet aphorisme n'est pas tout à fait exact: la crédibilité du recensement court un risque dans l'un ou l'autre cas. Nous ne devons pas attendre que des pressions politiques entraînent le changement, comme ce fut le cas aux États-Unis. Selon Brugioni (1981), le recensement de 1980 a été décrit comme: (traduction libre) "erroné" et comme: (traduction libre) "un exercice de futilité" par P.M. Klutznick, qui était alors secrétaire au Commerce.

Les objectifs actuels n'ont pas changé beaucoup depuis leur définition lors de l'élaboration du GRDSR. Fellegi (1967) a écrit que: (traduction libre) "Il est important d'entreprendre des travaux d'élaboration pour produire des outils généraux qui permettront de réaliser des économies afin de pouvoir traiter des quantités de données considérables et d'intégrer des éléments importants de normalisation". Dans le même document, il a préconisé l'économie: (traduction libre) "Nous devons nous assurer que la croissance à venir sera contrôlée et bien coordonnée et qu'elle sera réalisée par l'utilisation efficace des ressources financières et humaines".

Ce qui a changé depuis l'élaboration du GRDSR, c'est l'envergure et la portée du problème. Le volume de données aura augmenté de 500% d'ici 1991. On étudie la possibilité d'inclure les liens topologiques dans les fichiers de référence. De plus, une multiplicité de liens spatiaux sont évalués pour savoir s'ils devraient être inclus à la base de données afin d'étendre la gamme d'applications possibles.

Brackstone (1983) fait ressortir deux arguments très importants quand il écrit que: (traduction libre) "l'importance du recensement offre des occasions de recouvrir les investissements faits dans l'automatisation ... son importance cruciale implique que la technologie doit avoir été essayée et avoir fait ses preuves avant d'y être incorporée". Cela implique que le "crédit de relais" nécessaire pour effectuer la recherche initiale peut être récupéré pendant l'exploitation. Mais, en même temps, les systèmes doivent être prêts avant 1991 afin qu'on puisse en démontrer la fiabilité avant de les incorporer au processus de production du recensement.

Nous proposons donc:

- d'utiliser les résultats de la recherche effectuée pour le système TIGER afin d'éviter la répétition des efforts;
- de profiter du délai dont nous disposons pour effectuer un examen complet des systèmes;
- de ne pas essayer de tout faire en même temps, mais
- DE COMMENCER TOUT DE SUITE.

REMERCIEMENTS

Les auteurs voudraient remercier S. Witiuk et les autres membres de la Sous-division de la géocartographie ainsi que D.R. Bradley et ses subalternes de la Division de la géocartographie de leurs idées et commentaires. Nous tenons également à remercier les membres du Service d'exploitation du Centre de géocartographie qui ont su déchiffrer nos manuscrits et introduit le document dans les systèmes de traitement de texte de Statistique Canada.

BIBLIOGRAPHIE

Arnold, John, 1985. "A Police Application of AMF - The Geographic Resource Allocation Software System (GRASS)", **actes de la conférence de l'Urban and Regional Information Systems Association** de 1985, Ottawa.

Boisvenue, A., Bradley, R., Decarufel, M., Graves, R., Porter, A., Renaud, M., et Yan, J.,1983. "Geography Division Spatial Systems Review", rapport interne, Statistique Canada.

Brackstone, Gordon J., 1983. "The Impact of Technological Change on Census-taking", **Proceedings of the VIIIth Inter-American Statistical Conference**, Buenos Aires.

Bradley, D.R., 3 octobre 1985, note de service interne.

Brugioni, Dino A., 1983. "The Census: It Can Be Done More Accurately with Space-age Technology", **Photogrammetric Engineering and Remote Sensing**, volume 49, n° 9, p. 1337 - 1339.

Clayton, C., et Estes, J.E., 1980. "Image Analysis as a Check on Census Enumeration Accuracy", **Photogrammetric Engineering and Remote Sensing**, volume 46, n° 6, p. 757 - 764.

Collins, Stanley, H., Moon, George C., et Lehan, Timothy H., 1983. "Advances in Geographic Information Systems", **Les actes du sixième Symposium international sur la cartographie automatisée**, volume 1, Ottawa-Hull.

Fellegi, I.P., et Weldon, J.I., 1967. "Computer Methods of Geographical Coding and Retrieval of Data in the Dominion Bureau of Statistics, Canada", **Proceedings of the Social Statistics Section**, American Statistical Association, p. 53 - 57.

Gordon, Marvin, 1975. "Cartography for Census Purposes", **World Cartography**, volume 13, p. 3 - 18.

Ion, R.J., 1969. "The Geographic Basis of the DBS Geocoding System for Urban Areas: An Overview", note de service analytique et technique n° 3, Bureau fédéral de la statistique, Ottawa.

Marx, R.W., 1985. "Developing an Integrated Cartographic/Geographic Data Base for the United States Bureau of the Census", document rédigé à l'intention de Statistique Canada.

Nuttall, Gerry, et Korzenstein, Gary, 1985. "Evolution of a Regional Information System for the Metropolitan Toronto Planning Department", **actes de la conférence de l'Urban and Regional Information Systems Association de** 1985, Ottawa.

Parker, J.P., Yan, J., et Kidd, K., 1986. "Fichier principal de la région: un meilleur moyen de satisfaire les besoins des utilisateurs de données du recensement. Dans quelle mesure le programme devrait-il être élargi?" **Compte rendu de la conférence internationale sur la planification du recensement de** 1991, Statistique Canada.

Personnel de la Division de la géographie, 1982. "La géographie et le recensement du Canada de 1981 (série de la géographie)", document de travail n° 2 - GEO 82, Statistique Canada.

Peuquet, Donna J., 1984. "A Conceptual Framework and Comparison of Spatial Data Models", **Cartographica**, volume 21, n° 4, Toronto.

Peuquet, Donna J., 1983. "The Application of Artificial Intelligence Techniques to Very Large Geographic Data Bases", **Les actes du sixième Symposium international sur la cartographie automatisée**, volume 1, Ottawa-Hull.

Podehl, W.M., 1971. "An Introduction to the Generalized Tabulation Programs STATPAK and STATAPE and Their Language TARELA", document rédigé pour le Census Data Access Program Workshop, Statistique Canada.

Statistique Canada, juin 1972. "GRDSR: Renseignements par petits secteurs, manuel d'introduction, Ottawa.

Tomasi, S.G., 1986. "Soutien géographique pour le recensement décennal de 1990", **Compte rendu de la conférence internationale sur la planification du recensement de 1991**, Statistique Canada.

Tomlinson, R., 1984. "New Frontiers in Cartography", discours-programme prononcé à l'occasion du symposium international Spatial Data Handling de 1984, Zurich (Suisse).

White, Marvin, 1983. "Tribulations of Automated Cartography and How Mathematics Helps", **Les actes du sixième Symposium international sur la cartographie automatisée**, volume 1, Ottawa-Hull.

RÉSUMÉ DE LA DISCUSSION ÉTUDE COMPARATIVE DES PLANS AMÉRICAIN ET CANADIEN EN VUE DU RECENSEMENT DES ANNÉES 1990

SID WITIUK

DIVISION DES SERVICES ET DU DÉVELOPPEMENT INFORMATIQUES
STATISTIQUE CANADA

Introduction

Le premier exposé, de M. Silla Tomasi, nous a donné, sous une forme condensée, un aperçu très éloquent de ce qui est fait aux États-Unis à titre de préparation du recensement de 1990 en ce qui a trait au système de soutien géographique TIGER. Les trois exposés des représentants du Bureau ont donné une esquisse globale de ce qu'il est possible d'envisager, en termes de soutien géographique et cartographique, pour le recensement du Canada de 1991. Je me propose, pour ma part, de mettre en évidence certaines similitudes et certaines différences entre les deux recensements, d'y ajouter mes propres conclusions et de suggérer quelques points qu'il serait bon d'étudier plus en détail.

Voyons tout d'abord les **objectifs** poursuivis avec le système TIGER, les **caractéristiques du dossier commercial** et les **chances de réussite**.

1. TIGER

1.1 Objectifs

Les responsables du U.S. Bureau of the Census visent à atteindre les objectifs suivants: (1) réaliser une intégration totale des fichiers existants, (2) créer une **base de données** détaillées appuyée sur un fondement théorique solide et verticalement intégrée (c'est-à-dire que toute l'information géographique aussi bien que cartographique est ramenée à un niveau donné), (3) mettre sur pied un programme d'îlots adapté à l'ensemble du pays visant à faciliter la collecte, le dépouillement et la diffusion des données, et (4) élargir la portée du géocodage afin de réduire les coûts, par une utilisation accrue des méthodes d'envoi et de retour par la poste des questionnaires.

1.2 Dossier commercial

Qu'est-ce qui a amené les responsables du U.S. Bureau of the Census à proposer ces divers changements? En fait, ils ont présenté un dossier commercial solide. Nos homologues du Royaume-Uni voulaient savoir comment vendre un tel programme. C'est simple: il suffit de faire valoir que toutes les opérations s'effectuent **numériquement**, dans des délais **plus courts** que ce n'était le cas lors des recensements antérieurs, avec des **résultats améliorés** et pour un **coût moindre**; il est fort probable alors que l'autorisation et le budget nécessaires seront accordés sans difficulté. Une question demeure: "Le feront-ils vraiment?".

1.3 Chances de réussite

J'ai participé dernièrement, à titre de membre de l'équipe chargée de l'évaluation du système TIGER, aux travaux d'évaluation de ce système. Selon les conclusions de l'équipe, le défi que pose le système TIGER se situe plus au niveau de sa gestion qu'à un niveau technologique. TIGER doit assurer la coordination d'un volume considérable de données puisque près de 400,000 cartes (à raison de 5 à 6 exemplaires de chaque) doivent être produites dans un délai de 8 à 12 mois. C'est un travail énorme, mais l'équipe a jugé que du point de vue technologique cela est réalisable et devrait pouvoir être mené à bien.

L'équipe a souligné l'importance de certaines conditions préalables. **Il faut:** (1) que le système soit géré avec soin, (2) qu'il ne fasse pas l'objet de coupures, et (3) qu'il soit mis à l'essai assez longtemps à l'avance. Une des difficultés majeures qu'entraîne la transformation d'un système informatique aussi radicale que celle qui est envisagée aux États-Unis tient à ce que le produit final ne peut être connu avant que toutes les étapes intermédiaires aient été achevées. Pour le moment, les travaux en sont à la première étape et les deuxième et troisième étapes sont en préparation. Il est probable qu'au cours des dernières étapes, des problèmes vont surgir qui ont un lien avec les opérations des étapes antérieures. C'est pourquoi il est essentiel que TIGER soit mis à l'essai au plus tôt de façon que ces problèmes soient décelés assez rapidement. Il conviendrait également d'améliorer la coordination avec l'USGS. Bien que les relations de travail avec l'équipe de l'USGS soient exceptionnellement bonnes, la

coopération pourrait être améliorée en ce qui a trait à certains points de détail. Un effort a déjà été fait dans ce sens du côté américain, mais l'équipe a jugé qu'il y aurait encore à faire. En outre, le fait que les spécialistes en systèmes informatiques géographiques sont relativement rares, quel que soit le pays considéré, risque d'avoir une incidence sur la situation.

2. Système GCSS91

À partir des exposés faits par les spécialistes canadiens et des éléments du système en cause, connu sous le nom de Système géo-cartographique pour le recensement de 1991, système GCSS91 (acronyme que plusieurs aimeraient sans doute voir changé), reprenons les points que nous nous sommes proposés d'étudier: "Quels sont les objectifs du système, les caractéristiques du dossier commercial et les chances de réussite?".

2.1 Objectifs

Les objectifs visés avec le GCSS91 sont les suivants: (1) meilleure intégration des fichiers existants, (2) amélioration des structures pour les données géographiques et cartographiques **qui sont en relation**, (3) élaboration d'un programme d'îlots par étapes, sans doute adapté à la couverture du fichier principal de la région, (4) élargissement de la couverture du géocodage, pour répondre à la demande interne autant qu'externe, et (5) possibilité de production de cartes par étapes, également adaptée au fichier principal de la région, et disponibilité des fichiers cartographiques fournis par les diverses sources mentionnées par M. Yan.

Actuellement, on penche pour l'adoption d'une approche par étapes. La couverture intégrale du Canada ne sera donc probablement pas réalisée avant l'an 2001.

2.2 Dossier commercial

Bien entendu, si l'on opte pour une approche par étapes, le dossier commercial n'est plus aussi intéressant que dans le cas de l'approche directe. Par contre, le niveau de risque se trouve réduit de façon appréciable et le niveau d'automatisation est amélioré de façon proportionnelle. On obtient également une amélioration des délais et de la qualité (données plus cohérentes, notamment) et (grâce à une approche par étapes basée sur les conditions du marché) une réduction du coût unitaire, dans la mesure où le volume de données et le niveau d'automatisation augmentent. Les économies réalisées ne seront pas nécessairement apparentes du fait qu'elles seront utilisées pour améliorer le système tout au long du processus.

2.3 Chances de succès

L'élargissement et l'enrichissement du fichier principal de la région dépendent de la disponibilité des ressources plutôt qu'ils ne constituent un problème technologique important. Les questions qui se posent sont les suivantes: (1) "De combien de fonds dispose-t-on?", (2) "Qui participera au projet?", et (3) "Comment peut-on éviter une hausse du coût?". Au niveau de l'intégration, le problème se résume à une question de calendrier. Le programme d'îlots en est au stade de l'évaluation des coûts et profits et il faut encore attendre les résultats de cette analyse.

Dans l'état actuel des choses chacune des possibilités envisagées constitue, à mon avis, une démarche **raisonnable** qu'il convient d'entreprendre.

3. Comparaison des deux approches

Nous allons maintenant comparer les approches canadienne et américaine. De façon générale, les **analogies** se situent au niveau des objectifs. Les deux bureaux statistiques sont en train d'accroître le degré d'automatisation et d'intégration de leurs systèmes; ils s'orientent, à un rythme variable, vers des structures fondées sur la théorie et, à cet effet, mettent sur pied un programme d'îlots à portée nationale, élargissent la portée du géocodage, répertorient les adresses des individus sur support informatique et favorisent la cartographie numérique. Les **différences** se situent essentiellement au niveau des approches. Alors qu'aux États-Unis le Bureau of the Census vise l'automatisation et l'intégration **totales**, le Canada pour sa part privilégie: (1) une approche **par étapes**, (2) l'adoption de prototypes de sous-systèmes soigneusement élaborés, (3) l'adoption graduelle d'un système de production à l'échelle d'un territoire donné, et (4) l'élargissement graduel de la couverture par petites étapes.

Au Canada, les structures des fichiers sont modifiées en réponse à des décisions et des contraintes pragmatiques, alors que la structure du DIME/TIGER est basée sur la topologie et la théorie des graphiques. Parallèlement, dans le cas du programme d'îlots à portée nationale, l'approche envisagée par le Canada est une approche par étapes, basée sur les conditions du marché alors que le U.S. Bureau of the Census privilégie l'approche directe.

4. Impératifs pour le recensement de 1991

Nous venons de voir que si les deux recensements observés ont certains objectifs communs, ils diffèrent sur le plan de l'approche et des calendriers adoptés. Les deux bureaux continuent d'échanger le fruit de leur expérience propre et, en l'occurrence, l'ampleur des changements que le U.S. Bureau of the Census se propose de mettre en place est telle qu'il ne peut être qu'avantageux pour le Canada d'être en **deuxième position**. Le recensement canadien sera d'autant moins onéreux que le niveau de **risque sera moins grand**, non seulement sur le plan de l'élaboration de la théorie, des concepts et des systèmes, mais également en raison de la possibilité d'utiliser des ressources informatiques moins onéreuses et des données cartographiques numériques portant sur une longue période. La différence cruciale tient à ce que contrairement aux États-Unis, le Canada a un besoin résiduel de **crédit de relais** pour faciliter le passage à l'automatisation. Plus tôt le pas sera fait, plus tôt on tirera des profits des fichiers et plus le ratio avantages-coûts final sera grand. Une des grandes réussites du U.S. Bureau of the Census a été de modifier la courbe de consommation des ressources financières, c'est-à-dire d'abandonner le modèle classiquement adopté pour le recensement au profit de la **méthode de facturation avec majoration dégressive** appliquée à certains coûts, de façon à avancer la création de la base de données cartographiques. Au Canada, il conviendrait également d'imputer plus tôt les dépenses de recherche et d'élaboration afin d'assurer un recouvrement minimal des coûts et de façon que le développement de la base de données et la couverture soient régis par les forces du marché. En conséquence, il importe que la recherche et les études coûts-avantages soient entreprises au plus tôt.

5. Éléments-clés

Quels sont les principaux points qui doivent être étudiés, maintenant et au cours des prochains mois? Le premier consiste à **établir des objectifs réalistes** pour le recensement de 1991 et le second, à déterminer **dans quelle mesure** on doit, ou on peut, **prendre des risques**. Troisièmement, il faut établir des **calendriers** régissant la poursuite de divers objectifs et quatrièmement, déterminer de quelle manière on peut obtenir les fonds nécessaires pour maintenir et accroître la couverture des fichiers cartographiques numériques actuels. Cinquièmement, il convient de préserver et d'augmenter la part de **coopération** interne, nationale et internationale actuelle en vue de faciliter la mise sur pied de ces diverses initiatives. Sixièmement, la **décentralisation** est un objectif unanimement accepté mais dont la réalisation est différée en raison d'un certain nombre de contraintes. La décentralisation doit être étudiée du point de vue de son incidence possible sur le regroupement de la documentation (ce que le U.S. Bureau of the Census est en train de faire) et également du point de vue de l'établissement des tâches de collecte. Enfin, septièmement, il convient de déterminer de quelle manière on peut dériver une **mesure de l'efficacité** qui tienne compte de l'accroissement de l'automatisation. La comparaison schématique (voir la figure 1) des progrès réalisés des deux côtés de la frontière révèle que l'approche canadienne est beaucoup plus conservatrice que celle de nos homologues américains. De façon générale le chemin que nous avons parcouru l'a été par étapes et a été beaucoup moins ambitieux que dans le cas de nos voisins. À mon sens, il est temps que nous procédions à un changement majeur au niveau de la technologie de nos systèmes et de l'infrastructure géographique que nous utilisons. Cet effort devrait être assorti d'une mesure plus **circonscrite** en 1996. De cette manière, et sans entreprendre un changement aussi **drastique** que celui qui est envisagé par le U.S. Bureau of the Census, nous pourrions atteindre au recensement décennal de 2001 un niveau d'automatisation comparable à celui de nos voisins.

5.1 Perspective historique

Dans le domaine des systèmes informatiques spatiaux, on a fait de grands progrès technologiques depuis la fin des années 1960 et le début des années 1970 (voir la figure 2). Comme cela a été le cas pour beaucoup d'autres types de technologies durant la même période, le processus suivi comportait un certain nombre d'étapes bien définies, soit la **mise en marche**, la **prise de conscience** des possibilités et le bilan provisoire. (Ce schéma est valable pour toutes les applications géomatiques et non seulement pour le cas particulier du recensement.) Nous avons repris l'essai et avons eu à nouveau quelques problèmes mineurs. Finalement, nous avons atteint un seuil à partir duquel il sera plus facile d'avancer que cela n'a été le cas antérieurement.

5.2 Évolution observée

Le tableau 1 souligne qu'aux États-Unis l'adoption, à la fin des années 1960, de la méthode **d'envoi par la poste du questionnaire avec retour par la poste** a nécessité l'élaboration d'un **guide de codage des adresses**. Le **géocodage automatisé** qui en a découlé a, à son tour, nécessité le recours à la

topologie et conduit à l'établissement du système DIME. L'**extraction spécialisée** suppose l'existence de **coordonnées,** ce qui a conduit à l'élaboration du GRDSR et à la modification du DIME qui a pris la forme du GBF/DIME. Il semble que la production de cartes de dénombrement par ordinateur doit aboutir à la création d'une **base de données cartographiques numériques à portée nationale**.

Tableau 1. Évolution de la coparticipation

Envoi et retour par la poste	⟹	Guide de codage des adresses		
Géocodage automatisé	⟹	Topology	⟹	DIME
Extraction de données spatiales spécialisée	⟹	GRDSR	⟹	GBF/DIME
Cartes de dénombrement	⟹	CACM		
Base de données géographiques nationale	⟹	TIGER		

En guise de conclusion, je voudrais souligner que les **risques** encourus sont relativement **mineurs** par rapport aux **possibilités de changement** qui s'offrent à nous. Si l'on se fonde sur l'opinion exprimée récemment par le directeur du U.S. Bureau of the Census, M. Keane: (traduction libre) "le plus grand des risques est de ne prendre aucun risque", il semble sage d'opter pour ces risques mineurs et, ce faisant, de profiter des avantages offerts par une automatisation accrue.

Figure 1. Taux de changement technologique

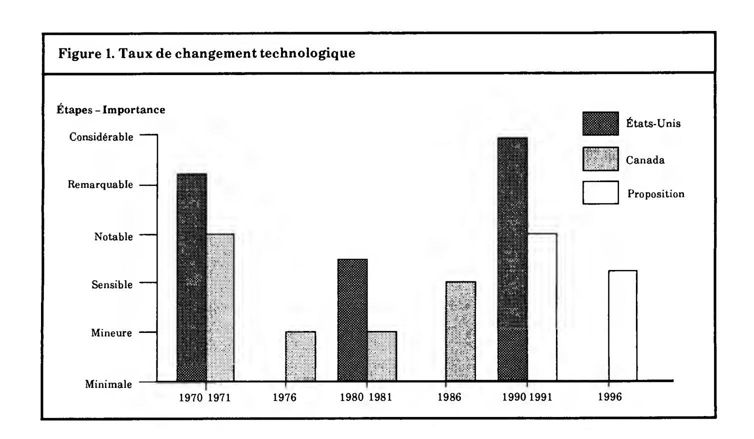
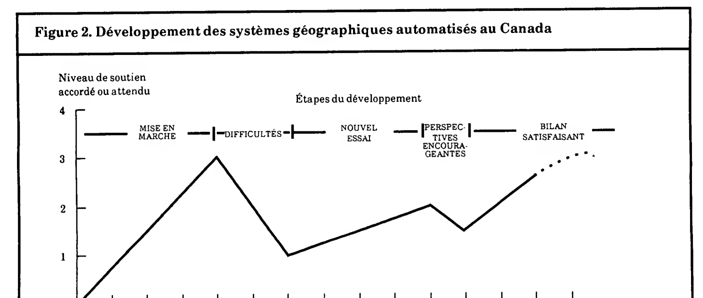


Figure 2. Développement des systèmes géographiques automatisés au Canada



Source: Taylor, D. R. F., rédacteur, Recent Trends in Geographic Information Processing in the National Capital Region, Department of Geography, Carleton University, Ottawa, mai 1978, p. 72.

Résumé de la période de questions

*L'exposé de **M. Tomasi** a été suivi d'une période de questions où plusieurs problèmes reliés aux recensements des années 1990 ont été soulevés. Deux questions en particulier portaient sur le Topologically Integrated Geographic Encoding and Referencing system (ou système TIGER).*

Collaboration interministérielle

Le premier sujet mentionné a fait l'objet d'une discussion générale portant sur les avantages potentiels des accords de collaboration entre les organismes chargés de la cartographie et ceux qui sont chargés des recensements.

Le U.S. Bureau of the Census (USBC) élabore le système TIGER pour 1990. Cependant, l'élaboration d'un tel système dans une période aussi courte exige des ressources et du matériel dont le USBC ne dispose pas pour le moment.

Un accord est engagé avec le U.S. Geological Survey (USGS) est au point mort. Une base initiale de données cartographiques sera produite par l'USGS. Au cours du recensement de 1990, l'USBC recueillera des renseignements qui seront utilisés pour mettre à jour la base de données cartographiques.

Quatre centres de cartographie de l'USGS travailleront exclusivement pour le programme du recensement américain. Cet accord de collaboration présente des avantages par rapport à l'utilisation du matériel et à l'emploi efficace de ressources spécialisées. Un autre avantage qui en découle est l'amélioration de l'uniformité des produits géographiques provenant de divers organismes.

Au Canada, Statistique Canada et Énergie, Mines et Ressources travailleront de concert pour créer une base de données numériques nationale d'ici 1990. Des programmes de collaboration existent aussi entre différents niveaux des administrations fédérale, provinciales et municipales afin de produire des cartes plus à jour. Cette mise à jour est généralement effectuée conjointement par les intéressés.

Un autre avantage possible d'une telle collaboration est la production de fichiers principaux de région (FPR) pour certaines régions urbaines de moins de 50,000 habitants. Cette augmentation du nombre de régions pour lesquelles un FPR est disponible représenterait une amélioration marquée et coûterait peu à chacun des intéressés.

Rôle des bureaux régionaux en ce qui a trait au système Tiger

Le second sujet qui a fait l'objet d'une discussion est le rôle joué par les bureaux régionaux (BR) dans l'élaboration du système TIGER.

Les BR ont joué un rôle important dans l'analyse qui a mené à l'élaboration du système TIGER. Ils ont été consultés pour déterminer les principaux défauts relevés lors du recensement de 1980 et ont fourni des renseignements précieux pour apporter des améliorations précises au système.

Le système TIGER est décentralisé en ce sens que les 12 régions de l'USBC sont chargées de la collecte des données de base, de la mise à jour des documents de base, d'une partie des travaux de numérisation et de la liaison avec les autorités locales et les responsables du programme des régions statistiques. Des spécialistes en géographie sont embauchés dans chaque BR afin d'entreprendre ces travaux.

SÉANCE: LE CONTENU DU RECENSEMENT DE 1991

Président: Ian Macredie
Division de l'analyse des enquêtes des ménages et du travail
Statistique Canada

Mercredi, le 9 octobre 1985



TABLE RONDE
PROCESSUS D'ÉLABORATION DU CONTENU DU RECENSEMENT DE 1991

HENRY KRIEGEL

BUREAU OF STATISTICS - AUSTRALIE

SUSAN MISKURA

BUREAU OF THE CENSUS - ÉTATS-UNIS

DAVID PEARCE

OFFICE OF POPULATION CENSUSES AND SURVEYS
ROYAUME-UNI

Résumé de la période de discussion

La première partie de la séance a consisté en une discussion qui portuit sur le processus d'élaboration du contenu du recensement de 1991 adopté par les organismes chargés du recensement en Australie, aux États-Unis et au Royaume-Uni.

1. Australie

M. Kriegel, représentant de l'Australian Bureau of Statistics, *a exposé le fondement légal de l'Austruliun Bureau of Statistics et présenté les grandes lignes du processus de consultation dans ce pays.*

L'Australian Bureau of Statistics Act, promulgué en 1975, accorde à l'Austrulian Bureau of Statistics (ABS) l'autorité nécessaire pour exercer son mandat et lui donne une autonomie accrue. En ce qui a trait à l'élaboration du contenu du recensement, la décision finale revient à l'Australian Bureau of Stutistics. En outre, la Loi de 1975 prévoit la création de l'Australian Statistics Advisory Council. Ce comité consultatif, dont les membres sont choisis à lu fois dans les secteurs public et privé, travaille de concert avec le personnel de l'ABS à l'élaboration du contenu du recensement.

Depuis le recensement de 1981, le processus d'élaboration du contenu vise à offrir à tous les utilisateurs des données du recensement la possibilité de faire des recommandutions.

Ce processus a été repris en 1986 parce que les responsables de l'Australian Bureau of Statistics avaient établi que le processus adopté pour le recensement de 1981 était satisfaisant et avait suscité un appui favorable du public vis-à-vis du recensement. M. Kriegel a alors brièvement présenté les principales caractéristiques du processus d'élaboration du contenu du recensement de 1986.

Une des premières étapes du processus d'élaboration du contenu incluait l'insertion dans les journaux de tout le pays de réclames invitant le public à faire des recommandations concernant les sujets qui devraient être couverts pur le recensement.

Puis, on a procédé à une série de consultations auprès des utilisateurs. Ces derniers étaient invités à faire des recommandations relativement aux questions du questionnaire de recensement. Ils pouvaient indifféremment proposer de modifier ou de supprimer des questions ou encore d'en ajouter de nouvelles. On a également demundé aux utilisateurs de se pencher sur le bien-fondé des sujets susceptibles d'être inclus dans les futurs questionnaires de recensement.

À partir des recommandutions recueillies, certains sujets ont ensuite été sélectionnés et une série d'essais sur le terrain a été entreprise en vue d'évuluer l'incidence des changements envisagés sur le furdeau de réponse et de déterminer si le recensement était le mode d'enquête approprié.

Une fois ces diverses évaluations effectuées, un document a été rédigé pour présenter les résultats préliminaires. Ce document fait état des recommandutions du Bureau relativement aux sujets à inclure dans le questionnaire de recensement ainsi que de ceux qui devaient être éliminés. Il y est également fait mention d'autres sources possibles qui permettraient de recueillir les données concernant certains sujets.

À la suite de cette consultation des utilisateurs, des recommandations ont été présentées à l'ISAC, puis soumises à l'approbation gouvernementale.

Il est apparu nécessaire de réviser le processus d'élaboration du contenu du recensement. D'après les consultations menées auprès des utilisateurs pour 1986, deux aspects en particulier devraient être améliorés: il s'agit de réduire les coûts du processus et la période nécessaire. Pour ce faire, on pourrait, notamment, utiliser divers rapports d'enquête mensuels pour inviter les intéressés à proposer des sujets à inclare dans le questionnaire de recensement. Une liste de sujets possibles pourrait être donnée sur la couverture arrière de ces rapports d'enquête avec une note invitant les utilisateurs à faire part de leurs recommandations.

Quelle que soit la méthode retenue pour l'élaboration du contenu du recensement de 1991, l'expérience des deux derniers recensements a révélé que l'implication des principaux utilisateurs faisant partie du public est essentielle.

2. États-Unis

Mad. Miskura, représentante de l'U.S. Bureau of the Census, *a présenté le processus de consultation adopté aux États-Unis.*

Le processus et le calendrier d'élaboration da contena sont établis essentiellement en fonction des contraintes imposées au Bureau par le gouvernement: les sujets couverts par le recensement de 1990 doivent être connus le 1er avril 1987 et le libellé des questions doit être déterminé un an plus tard.

Pour déterminer le contenu du recensement de 1990 on a commencé par analyser celui du recensement de 1980. Cette analyse comportait une évaluation de la qaalité des données et de l'utilisation de ces dernières par divers utilisateurs. Ceux-ci ont été invités à participer à des forums conçus spécifiquement pour mettre en évidence les besoins de ce point de vue. Ces forums ont pris la forme, principalement, d'assemblées publiques locales, organisées par le Census Bureau avec la collaboration d'organismes relevant du gouvernement ou d'organismes locaux. Ils ont fourni l'occasion à une grande variété d'utilisateurs du secteur privé aussi bien que public de formuler leur opinion relativement à la précision des données et de proposer des modifications ou des innovations pour le prochain recensement. Un foram de ce type, aa moins, a été tenu dans chaque état. D'autres réunions, des conférences et les démarches faites par le Bureau constituent également les principales sources de renseignements à partir desquels on a élaboré le contenu du recensement.

Pour déterminer les besoins de données des organismes fédéraux, le Census Bureau a constitué un conseil qui regroupe des représentants de ces organismes, sous la forme de dix groupes de travail, appelés Interagency Working Groups ou IWG, et également sous la forme de l'Office of Management and Budget's Federal Agency Council on the Decennial Census. Les IWG sont composés d'experts représentant quelque 80 organismes fédéraux. Les principaux objectifs de ces groupes de travail sont les suivants: obtenir de l'information concernant l'utilisation des données de recensement, tant d'un point de vue législatif qu'informatique, déterminer parmi les données de 1980 les éléments et les totalisations qu'il n'est pas nécessaire de reprendre en 1990, exposer leurs domaines d'intérêt et leurs préoccupations en suggérant les modifications appropriées au libellé des questions ou de nouvelles catégories de réponses possibles pour le recensement de 1990 et préciser les niveaux géographiques pour lesquels des données devraient être produites. Les IWG oni présenté les besoins de données par ordre de priorité: indispensable, nécessaire ou soahaitable, en spécifiant si les renseignements pouvaient être recueillis auprès d'un échantillon de la population.

La sélection des sujets couverts par le recensement est régie par un principe général. Le recensement doit servir uniquement à produire les données requises pour répondre à des besoins clairement établis du public, pour remplir les obligations fixées par la loi ou pour mener à bien les programmes gouvernementaux. D'autres critères sont également pris en considération comme le besoin de données portant sur de petites régions géographiques ou sur des groupes de petite taille et fort dispersés, la possibilité d'obtenir de telles données de sources autres que le recensement, la faisabilité d'une formulation appropriée de la question et le coût lié à l'insertion du nouveau sujet.

Les spécialistes du Census Bureau ont alors procédé à une première sélection des nouvelles questions proposées et ont mis sur pied un programme d'essai. La composante principale de ce programme est connue sous le nom de National Content Test. Ces essais visaient à évaluer la fiabilité des données recaeillies ainsi

qu'à déterminer dans quelle mesure les recensés pouvaient répondre aux questions posées et acceptaient de le faire.

Un des objectifs essentiels du recensement de 1990 est de trouver un équilibre entre les besoins croissants de nouvelles données et la nécessité de ne pas alourdir le fardeau de réponse, c'est-à-dire de ne pas allonger indûment le temps nécessaire pour remplir le questionnaire.

Soucieux de maintenir le fardeau de réponse à un niveau raisonnable, le Bureau est en train d'étudier la faisabilité d'une réduction du nombre des questions. La façon la plus simple d'y arriver consiste à supprimer certaines questions ou à éliminer certains sujets du questionnaire destiné à l'ensemble de la population. Une autre solution possible consisterait à utiliser plusieurs questionnaires-échantillon ce qui permettrait de recueillir de nouvelles données sans augmenter le temps moyen nécessaire pour répondre au questionnaire du recensement. Ces stratégies, ainsi que d'autres, permettraient d'atteindre un équilibre entre le besoin croissant de données et la nécessité de ne pas accroître le fardeau de réponse.

3. Royaume-Uni

M. Pearce, représentant du Royaume-Uni, *a exposé le processus de consultation adopté au Royaume-Uni en présentant les parties visées par la consultation, en décrivant la manière dont celle-ci doit être faite et la période opportune pour la faire et, enfin, en spécifiant le poids que l'on doit accorder aux suggestions des utilisateurs au moment de l'élaboration du contenu du recensement.*

Il y a plusieurs types d'utilisateurs qui doivent être consultés. Le gouvernement central a besoin de données pour allouer les fonds, les gouvernements locaux ont besoin de données portant sur des régions géographiques de petite taille, les universitaires ont besoin de données pour leurs recherches et, enfin, le monde des affaires a besoin des données du recensement pour élaborer ses stratégies commerciales. Il ne fait aucun doute que les données du recensement sont utiles à divers groupes des secteurs public et privé. Si l'on veut être en mesure de satisfaire ces besoins diversifiés, il est essentiel que la consultation des utilisateurs intéressés soit faite au moyen d'un programme structuré.

Il y a deux niveaux possibles de consultation: l'un consiste à consulter la masse des utilisateurs de façon informelle, à l'occasion de réunions spéciales, et l'autre correspond à une consultation plus formelle et davantage structurée, menée au moyen d'une série de réunions avec les représentants d'autres organismes gouvernementaux.

Le premier type de consultation peut prendre une variété de formes; il peut s'agir de réunions spéciales ou encore d'une série de réunions informelles tenues dans tout le pays. Cette façon de faire peut s'appliquer à l'ensemble des utilisateurs, notamment aux gouvernements locaux. En 1981, ce type de réunions a eu un effet réel puisque cela a conduit à l'inclusion de nouvelles questions telles que celles portant sur le logement, sur l'adresse cinq ans auparavant et sur l'origine ethnique. Au chapitre des consultations informelles, citons également la série de publications "Monitor Series" qui est distribuée gratuitement à quelque 5,000 personnes susceptibles d'être intéressées par les données du recensement.

Le processus de consultation formelle est davantage structuré. En 1981, l'Office of Population Censuses and Surveys (OPCS) a organisé une série de réunions avec les représentants d'autres ministères pour discuter de sujets déterminés et a mis sur pied un comité consultatif regroupant notamment les directeurs des services statistiques des ministères, les responsables des gouvernements locaux et les chercheurs universitaires les plus intéressés au sujet.

À la suite de ces consultations, l'OPCS a créé un comité spécial et rédigé un document de travail traitant du contenu du recensement, qui incluait un certain nombre de recommandations et proposait d'autres sources possibles de données. Ce document a été distribué aux intéressés avant d'être présenté au Parlement.

Qu'il s'agisse de consultation formelle ou informelle, en ce qui a trait aux questions dont il convient de recommander l'inclusion dans le questionnaire du recensement, la décision finale revient à l'OPCS.

Un dernier point reste à considérer; il s'agit des délais de la consultation. Du fait que les questions doivent figurer dans le questionnaire qui est soumis à l'approbation du Parlement plusieurs années avant le recensement, le processus de consultation doit être entamé quatre ou cinq ans avant l'année du recensement. Par conséquent, il faut réduire la durée du processus de consultation de façon à déterminer au plus tôt les particularités des sujets qui devront être inclus dans le questionnaire de recensement.

TABLE RONDE
PROPOSITIONS RELATIVES AU CONTENU DU RECENSEMENT DE 1991

JOHN KETTLE

FUTURESEARCH PUBLISHING INC.

FRANK CLAYTON

CLAYTON RESEARCH ASSOCIATES

NOAH MELTZ

UNIVERSITÉ DE TORONTO

Discussion

La deuxième partie de la séance a consisté en la présentation, par trois principaux utilisateurs des données du recensement, de propositions relatives au contenu du recensement de 1991.

1. Futuresearch Publishing Inc.

M. Kettle, porte-parole de Futuresearch Publishing Inc., a fait quelques suggestions concernant les données qu'il conviendrait de recueillir en 1991. Ces suggestions sont formulées non seulement parce qu'elles reflètent les besoins intrinsèques du groupe que représente leur auteur, et ceux d'autres utilisateurs probablement, mais également parce qu'elles présentent en outre l'intérêt de permettre certaines corrélations avec des données régulièrement recueillies dans un recensement, comme l'âge, le sexe, l'instruction, le revenu et la profession. Ces propositions ont été regroupées sous trois rubriques.

Santé

Alors que des données très valables sont disponibles concernant l'utilisation globale des services médicaux, il n'existe aucune donnée sur l'utilisation individuelle des hôpitaux et autres services médicaux. Il serait très utile d'avoir de telles données sur une base actuelle (pour les 12 derniers mois) et cumulative (pour la durée de vie). Ces données pourraient inclure le nombre de visites chez le médecin, le nombre de séjours à l'hôpital, le total des congés de maladie, le type de maladie, etc. Même si une seule de ces variables était incluse, ce serait déjà appréciable.

Éducation

Il serait extrêmement avantageux de disposer de données actuelles sur les activités éducatives et de données cumulatives sur le niveau de scolarité qui soient de meilleure qualité. L'éducation est indispensable pour établir les prévisions relatives à d'autres variables telles que l'activité, le taux de fécondité chez les femmes, etc., et le fait de disposer de données de meilleure qualité permettrait d'améliorer ces prévisions. Les données recueillies devraient englober les études à temps partiel et à plein temps, la formation et le recyclage. Dans le cas des personnes qui ont suivi des études postsecondaires, il serait bon de connaître le ou les domaines d'études, notamment pour établir les prévisions relatives à la profession.

Travail

Les suggestions sont nombreuses à ce chapitre, même s'il est difficile d'établir une nette distinction entre les objectifs spécifiques du recensement et de l'enquête sur la population active.

- **Lieu de travail.** Le but visé est d'avoir une meilleure image de l'augmentation rapide du nombre de travailleurs dans les régions rurales non agricoles. Le classement recoupé des données sur le lieu de travail selon le lieu de résidence devrait permettre d'éclaircir certains mystères.

- **Travail à la maison.** On estime à environ 10% le nombre des personnes qui travaillaient à la maison (ou à partir de chez elles). Compte tenu de l'état d'avancement actuel de la technologie, on estime à 40% le nombre de personnes qui pourraient opter pour ce mode de travail. Par ailleurs, on prévoit qu'en 1995, 20% de la population **travaillera** à la maison. Aucune enquête actuelle ne comporte de question sur ce sujet.

- **Gestionnaire, selon le niveau.** Il y a actuellement environ un million de gestionnaires dans la population active, soit deux fois plus que le nombre de travailleurs dans le secteur primaire, que le nombre de travailleurs de la construction et que le nombre de travailleurs dans le secteur du transport, mais nous ne disposons d'aucune donnée sur le niveau de ces gestionnaires. Du fait que cette catégorie connaît de nombreux changements et que l'évolution en cours va probablement s'accentuer au cours des 10 prochaines années, il importe d'avoir des données de meilleure qualité. Une répartition des données selon les catégories "cadres supérieurs", "cadres intermédiaires" et "superviseurs" serait déjà appréciable. La corrélation avec la variable "sexe" permettrait de mettre en évidence des changements qui vont avoir une certaine importance.

- **Propriétaires.** Les remarques que nous venons de faire relativement aux gestionnaires s'appliquent également aux données concernant les propriétaires: celles-ci sont insuffisantes. La corrélation avec les variables "âge" et "sexe" révélerait des tendances qui vont prendre de plus en plus d'importance.

- **Travailleurs du secteur relié à l'information.** Il risque d'être impossible de poser des questions sur la nature du travail effectué qui appellent des réponses conséquentes. Or, actuellement, près de la moitié de la population active occupée fait un travail qui est relié à l'information, bien que cette catégorie de travailleurs ne soit prise en considération par aucune des enquêtes ni par le recensement.

- **Petites entreprises.** Le même problème se pose, et probablement de façon plus accentuée, lorsqu'il s'agit de déterminer le nombre de personnes qui travaillent pour de petites entreprises (les entreprises de moins de 20 employés constituent une catégorie intéressante de ce point de vue). Il était possible, à tout le moins, de produire des estimations de cette sous-population jusqu'à ce que Statistique Canada publie la nouvelle série relative à l'emploi, au printemps de 1983; maintenant ce n'est plus possible. Or, la majorité des nouveaux emplois s'ouvrent dans de petites entreprises.

- **Économie parallèle.** De toute évidence, il est impensable de demander aux gens de déclarer le travail effectué sans facture ou au noir; toutefois, il peut être utile de soulever la question à l'occasion de ce forum au cas où quelqu'un pourrait proposer une façon de recueillir des renseignements utiles sur ce type de travail dont l'importance est considérable et va grandissant.

- **Loisirs.** Les activités de loisir occupent actuellement une part plus importante du temps d'éveil que le travail formel. Nous sommes prêts à consacrer nos efforts à l'analyse de ce qui se passe au travail; nous devrions également nous intéresser au temps consacré aux loisirs. La répartition du temps occupé à des activités autres que le travail peut se faire selon un grand nombre de catégories telles que: soins donnés aux enfants, travaux ménagers, sport, exercice physique, loisirs à domicile, télévision, lecture, jardinage, artisanat, temps mort, activités culturelles à l'extérieur, distractions, études, formation.

2. Clayton Research Associates

Selon **M. Clayton, porte-parole de Clayton Research Associates,** il est essentiel d'obtenir des données sur le logement. Les utilisateurs des données du recensement du Canada s'attendent, depuis 1961, à ce que chaque recensement décennal inclue une série de questions sur le logement portant notamment sur le mode d'occupation, le type de logement, les coûts, l'âge, le type de chauffage et le nombre de pièces. Les données recueillies à l'occasion des recensements quinquennaux sont habituellement beaucoup moins diversifiées. En comparaison, la variété des questions posées au recensement de 1986 relativement au logement est une nouveauté appréciée.

Raison d'être des questions sur le logement dans le questionnaire du recensement

Les données sur le logement tirées du recensement sont largement utilisées pour l'élaboration de politiques et à des fins de commercialisation:

- **Politiques relatives aux logements sociaux.** Les responsables de politiques gouvernementales chargés d'analyser les problèmes des Canadiens au chapitre du logement ont besoin de données illustrant le rapport entre le coût du logement et le revenu pour tous les ménages, selon le type et les caractéristiques du logement.

- **Politiques visant à maintenir ou à améliorer la qualité du stock immobilier.** Depuis le début des années 1970, les pouvoirs publics attachent de plus en plus d'importance au maintien et à l'amélioration de la qualité du stock de logements.

- **Politiques relatives à l'énergie.** Vers la fin des années 1970, les pouvoirs publics ont porté un intérêt croissant aux types d'installation et de combustible utilisés pour chauffer ou climatiser les logements.

- **Politiques visant un usage plus large du stock immobilier existant.** Dans les vieux quartiers des villes canadiennes, il est possible d'augmenter le nombre de logements car beaucoup de maisons de grande taille sont occupées par une seule personne.

- Pour pouvoir effectuer des analyses de marché valables, on a besoin de données exactes sur les caractéristiques des ménages, sur leur revenu ainsi que sur les conditions du marché immobilier, du fait notamment que la demande pour les nouveaux logements devient de plus en plus morcelée.

Les diverses utilisations que nous venons d'énumérer ont un point commun; elles soulignent la nécessité d'une base de données détaillées couplant les variables propres au logement et les caractéristiques démographiques et socio-économiques de la population.

Ces données ne peuvent-elles être recueillies au moyen d'enquêtes-échantillon? Pourquoi, par exemple, l'enquête sur le revenu des ménages et l'équipement ménager, menée par Statistique Canada, ne permet-elle pas de répondre à ce besoin? S'il est vrai que les enquêtes-échantillon fournissent des données utiles, il demeure qu'il y a un besoin de données sur le logement obtenues par recensement. Ce besoin s'appuie sur trois raisons:

(1) Nécessité de données repères.

Comme c'est le cas pour d'autres données recueillies dans le cadre du recensement du Canada, notamment le chiffre total de population et les caractéristiques de celle-ci, il est indispensable d'avoir périodiquement des chiffres précis en ce qui a trait à certaines données clé sur le logement telles que le nombre total de logements (occupés et inoccupés) selon le mode d'occupation et le type. Les données sur les logements prêts à être occupés ne constituent pas un substitut valable car le stock de logements existant subit des modifications constantes. Par ailleurs, du fait que beaucoup des utilisations des données sur le logement ont trait à l'évolution du stock immobilier et de ses caractéristiques, les enquêtes-échantillon sont inadéquates car les variations mises en évidence englobent les changements actuels mais aussi l'erreur d'échantillonnage. Ainsi, on ne peut connaître avec précision l'évolution au niveau de la répartition du stock de logements selon le mode d'occupation depuis 1981 à cause de l'erreur d'échantillonnage qui touche les estimations disponibles (basées sur les résultats des enquêtes-échantillon).

(2) Disponibilité de données régionales.

Cet aspect du recensement est probablement son plus grand avantage du point de vue des données sur le logement. Au Canada, le recensement est la seule source qui offre des données sur le logement (ainsi qu'une série détaillée intégrant des données sur le logement et sur les caractéristiques démographiques et socio-économiques) pour des régions géographiques de petite taille.

En quoi cela est-il essentiel?

a) Les marchés immobiliers, de par leur nature même, sont très localisés (par exemple, Ottawa ou Halifax) et chacun de ces marchés locaux est lui-même constitué de sous-marchés. Cela tient à ce que le logement est une marchandise unique qui n'est pas mobile et qui a une longue durée.

b) Les responsables de politiques gouvernementales ont besoin de données régionales notamment pour repérer les secteurs qui se caractérisent par une forte concentration de logements en mauvais état.

Le recensement du Canada offre également l'avantage de permettre la sélection de limites géographiques spécifiques du fait que les données recueillies sont géocodées.

(3) Manque de substitut valable aux données du recensement.

Les enquêtes-échantillon, comme l'enquête sur les dépenses des familles, l'enquête sur les finances des consommateurs et, plus particulièrement, l'enquête sur le revenu des ménages et l'équipement ménager fournissent des données utiles sur le logement et sur des variables liées au logement. Toutefois, ces enquêtes n'offrent pas les avantages du recensement, soit des données chronologiques, des données très fiables et des données régionales. Les enquêtes présentent de sérieuses limites lorsqu'il s'agit d'établir des estimations pour les provinces les moins peuplées, sans parler des sous-régions de petite taille.

Pertinence des données sur le logement recueillies dans le cadre du recensement

En 1981, le recensement a permis de recueillir une variété de données sur le logement, concernant notamment:

- le stock de logements occupés selon le mode d'occupation et le type de logement,
- les caractéristiques des logements, à savoir l'âge, le type de chauffage et de combustible, le nombre de pièces, le nombre de salles de bains et l'état du logement,
- les coûts du logement (pour les propriétaires aussi bien que pour les locataires) et la valeur des logements occupés par leur propriétaire.

Dans l'ensemble, cette variété de données permet de satisfaire les besoins de la plupart des utilisateurs.

Quelques questions dans le questionnaire du recensement de 1981 pourraient, réflexion faite, paraître d'utilité marginale. Il s'agit notamment de la question sur le type de combustible utilisé pour chauffer l'eau et de celle sur le nombre de salles de bains complètes et de demi-salles de bains. On peut aussi se demander pourquoi la question sur la durée d'occupation a été incluse parmi les questions relatives au logement, en 1981.

Besoins en données sur le logement tirées du recensement de 1991

En ce qui a trait à l'habitation, la situation est en pleine évolution. Au cours des années 1990, le niveau de la construction de nouveaux bâtiments résidentiels va baisser de manière prononcée par suite de l'incidence des forces démographiques. Ce déclin, de même que le vieillissement du stock immobilier existant, pousseront les responsables des secteurs public et commercial à porter une attention de plus en plus marquée au maintien et à l'amélioration du stock existant.

On note actuellement des lacunes importantes dans les statistiques relatives à la rénovation des logements résidentiels et cette situation ne peut que s'accentuer. Il conviendrait d'étudier soigneusement des façons d'y remédier à l'occasion du recensement de 1991. On pourrait, par exemple, recueillir des données sur l'importance et le type des rénovations entreprises par les propriétaires de logements au cours des 12 derniers mois, sur les montants consacrés à ces rénovations (par activité) par les propriétaires de logements au cours des 12 derniers mois ou encore sur certains aspects du stock immobilier (par exemple, le nombre de fenêtres, l'existence d'un garage).

Risque de conflit entre la continuité chronologique et l'insertion de nouveaux sujets à l'occasion du recensement de 1991

La question de la continuité des données recueillies dans le cadre des recensements est une préoccupation sérieuse. Cependant, pour autant que les questions relatives au logement qui seront posées à l'occasion du recensement de 1991 auront la même portée que celles du recensement de 1981, il ne devrait pas y avoir de problème véritable du point de vue de la cohérence des données chronologiques. La fiabilité des données existantes (par exemple, les données sur le type de logement recueillies en 1981) constitue, plus que la continuité, une préoccupation majeure.

Équilibrage de la demande illimitée des utilisateurs pour des données sur le logement et de l'espace limité alloué à ce sujet dans le questionnaire

Là encore, si la portée du recensement de 1991 est la même que celle du recensement de 1981, Statistique Canada n'aura pas à faire face à des pressions considérables visant l'inclusion de nombreuses nouvelles questions sur le logement dans le questionnaire de recensement.

En plus d'être la seule source adéquate de données sur le logement, le recensement est

également la méthode de collecte la plus rentable pour ce type de données.

3. Université de Toronto

M. Meltz, de l'Université de Toronto, a présenté ses suggestions concernant les améliorations à apporter au contenu du recensement de 1991. Son exposé, qui traite des données relatives au marché du travail, s'articule autour de quatre points principaux: les conditions d'utilisation du recensement comme source de données sur le marché du travail, les domaines d'intérêt probables, relativement au marché du travail, dans les années 1990, les données sur le marché du travail tirées du recensement de 1981 et les commentaires relatifs au contenu du recensement de 1991.

Conditions d'utilisation du recensement comme source de données sur le marché du travail

Avant de faire des propositions concernant le recensement de 1991, il convient de rappeler quel est l'objectif de tout recensement décennal, autrement dit de préciser ce qui peut ou ne peut pas être fait. Le recensement décennal est une source de données détaillées portant sur diverses caractéristiques de la population canadienne. En ce qui a trait au marché du travail, le recensement fournit de manière exclusive des statistiques socio-économiques en rapport avec le revenu, l'emploi et le chômage de la population active classée selon la profession, ce qui, en 1981, donnait quelque 500 professions correspondant à environ 300 industries, c'est-à-dire établissements industriels, commerciaux ou de services. Ces caractéristiques socio-économiques sont: l'âge, le sexe, le niveau et le type de scolarité, l'origine ethnique, la religion, la langue, le pays d'origine, la période d'immigration, le lieu de résidence, le lieu de travail, etc. Aucune autre enquête n'offre des données aussi détaillées sur une aussi grande diversité de professions et de régions géographiques que le recensement (Meltz, 1982).

Toutefois, l'étendue de la couverture de la population et le détail des données en ce qui a trait à certains sujets entraînent certaines contraintes. L'une d'elles est que le questionnaire couvre un nombre nettement moins important d'aspects du marché du travail que l'enquête permanente correspondante, soit l'enquête sur la population active (EPA). Cette dernière traite beaucoup plus en détail les sujets tels que le nombre d'heures travaillées, la recherche d'emploi et le chômage. Par contre, les données du recensement sur l'éducation, la formation et le revenu sont plus diversifiées. Bien que les deux types d'enquête aient des points communs, leurs buts divergent sensiblement. L'EPA vise à brosser un tableau détaillé des antécédents de la population active relativement à l'emploi et au chômage. Toutefois, elle ne fournit pas de données sur des professions ou des industries spécifiques ni sur de petites régions du pays. Le recensement constitue une source de référence particulièrement détaillée.

Domaines d'intérêt probables, relativement au marché du travail, dans les années 1990

Avant de proposer des modifications (ou des ajouts) au contenu du recensement, il convient de préciser en quoi cela est nécessaire. Ces changements sont commandés, notamment, par l'évolution prévue des domaines d'intérêt au chapitre du marché du travail, dans les années 1990. Quels aspects devront être couverts par le prochain recensement afin de répondre à la demande de données des années à venir? En voici quelques-uns: la structure démographique de l'ensemble de la population et de la population active de façon générale et par rapport aux diverses professions et industries, les pensions et la situation vis-à-vis de l'activité d'une population comptant de plus en plus de personnes âgées, les autres revenus ne provenant pas d'un emploi et les avantages sociaux liés au travail, l'égalité d'accès à l'emploi, la parité de salaire à travail égal, l'augmentation du nombre de travailleurs à temps partiel, la rémunération des travailleurs des secteurs privé et public respectivement, le chômage et l'immigration.

Données sur le marché du travail tirées du recensement de 1981

Le recensement permet d'avoir des données sur l'offre de main-d'oeuvre et sur les résultats du jeu de l'offre et de la demande de main-d'oeuvre, c'est-à-dire sur l'emploi, sur la rétribution du travail (traitements, salaires et avantages sociaux) et sur le chômage. Le recensement ne peut mesurer la demande non satisfaite de main-d'oeuvre.

Le recensement est une source essentielle de données sur la structure démographique et sert de point central de référence pour de

nombreuses enquêtes. Il fournit également des données sur la période d'immigration et le pays d'origine de même que sur les pensions et sur la situation vis-à-vis de l'activité, par groupe d'âge.

En ce qui a trait à l'offre de main-d'oeuvre, le recensement de 1981 donne également des renseignements concernant le niveau de scolarité et la formation. Il ne fournit pas un tableau détaillé du niveau de scolarité et de la formation de la population, mais il donne une idée du niveau général et du type de résultats atteints.

Au chapitre de l'incidence de l'offre et de la demande sur le marché du travail, le recensement fournit des données sur certains aspects de l'emploi, du revenu et du chômage. Ces données peuvent ensuite être recoupées avec les données relatives aux caractéristiques socio-économiques.

Dans la perspective des sujets d'intérêt dans les années 1990, il y a plusieurs lacunes qu'il conviendrait de combler. Ainsi, il faudrait recueillir des données supplémentaires sur les heures de travail et les taux de rémunération, du point de vue du travail à temps partiel, de l'égalité d'accès à l'emploi et de la parité de salaire à travail égal. Contrairement à l'enquête sur la population active, le recensement ne fait pas la distinction entre les heures travaillées dans le cadre de l'emploi principal et celles qui correspondent à un ou des emplois secondaires (questions 39 et 44 du questionnaire de recensement et questions 18, 76 et 77 du questionnaire de l'EPA). Les données sur les salaires et traitements tirées du recensement sont attribuées à un emploi alors que les revenus correspondants pour l'année visée peuvent provenir de plusieurs emplois. De plus, et cela est plus grave, la profession déclarée n'est pas nécessairement celle dont le recensé a tiré la plus grande part de ses gains.

Un autre aspect lié à ces deux mêmes sujets est le taux de rémunération, c'est-à-dire la rétribution touchée pour une heure ou une période de temps déterminée. Les données du recensement de 1981 ne permettent pas de calculer le taux de rémunération, mesure de base de l'analyse de l'économie du travail. La question sur le nombre de semaines travaillées (question 45) porte sur l'année 1980 alors que celle sur les heures travaillées (question 39 a)) a trait à la première semaine de juin 1981. Du fait que les périodes visées sont différentes et que les traitements et salaires pourraient provenir de plusieurs sources, il est difficile d'estimer le taux de rémunération.

Les données du recensement permettent seulement une comparaison partielle des taux de rémunération et des gains réalisés dans les secteurs public et privé. Comme il a été précisé plus haut, il n'est pas possible de calculer les taux de rémunération, pas plus qu'il n'est possible d'évaluer l'emprise des syndicats. Enfin, si la classification des industries comporte une catégorie relative à l'ensemble des emplois fournis par le gouvernement, elle n'identifie pas de façon distincte les organismes qui sont la propriété de ce dernier.

Alors que le recensement donne une répartition détaillée du revenu selon la source, il ne fournit pas de données sur les divers avantages sociaux tels que les congés payés, les jours fériés, la part de l'employeur aux régimes d'assurance-maladie, les régimes d'épargne, etc.

Commentaires relatifs au contenu du recensement de 1991

Compte tenu de ce qui précède, voici quelques commentaires relativement au contenu du recensement de 1991. Ces suggestions sont basées sur l'hypothèse que l'objectif du recensement est toujours de fournir des données de référence détaillées et que le nombre des questions est limité. Tout d'abord, il conviendrait d'élargir le champ des questions de 1981 sur les heures travaillées de façon à établir le nombre d'heures qui correspond à l'emploi principal et aux emplois secondaires (question 39 a)). La même remarque s'applique aux questions sur la catégorie de travailleurs (question 44), sur le nombre de semaines travaillées (question 45) et sur le montant perçu au titre des salaires et traitements (question 46).

Un changement de même ordre doit être apporté aux questions 41 et 43. Nous avons toujours considéré que les tableaux croisés regroupant les données sur la profession et celles sur les salaires et traitements étaient parfaitement cohérents, c'est-à-dire que les gains s'appliquaient effectivement à la profession correspondante dans le tableau. Cela était vrai jusqu'à 1961 inclusivement. Or, d'après le libellé de la question 43 (et, par analogie, de la question 41 relative à l'industrie) il s'agit de la profession durant la semaine précédant le 3 juin 1981 qui peut être

différente de la profession dont les revenus ont été tirés en 1980 (question 46 a)). Il est fort probable qu'il s'agisse de la même profession dans la plupart des cas où les personnes ont travaillé au cours de l'année. Les mêmes possibilités s'appliquent aux personnes n'ayant pas travaillé durant la dernière semaine de mai 1981. Néanmoins, il faudrait préciser les questions 41 et 43 de façon à établir si la profession et l'industrie déclarées s'appliquent bien à la rémunération touchée en 1980. Faute de quoi, il faudra prévoir une question sur le principal emploi et l'industrie correspondante en 1980.

Un autre ajout, mineur mais très utile, consisterait à préciser à la question 45 du questionnaire de recensement que le temps partiel correspond à moins de 30 heures de travail. Cette précision est incluse dans la question 39 d), mais non dans la question 45. On pourrait, tout aussi bien, préciser aux questions 39 d) et 45 que le travail à plein temps correspond à 30 heures ou plus, comme cela est indiqué dans l'enquête sur la population active.

Soulignons ici une légère différence dans le libellé des questions du questionnaire de l'EPA et de celui du recensement de 1981. La question 39 e) du questionnaire du recensement demande s'il y a une raison qui vous ait empêché de commencer à travailler la semaine dernière. La question 64 du questionnaire de l'EPA demande s'il y a une raison pour laquelle ... n'a pu prendre un emploi la semaine dernière. Pour une meilleure compatibilité des données, il faudrait utiliser le terme "prendre" au lieu du terme "commencer".

Il convient de s'arrêter à un autre aspect, à savoir la durée d'emploi pour le même employeur. L'inclusion d'une question similaire à la question 73 de l'EPA portant sur la date à laquelle la personne a commencé à travailler pour cet employeur ajouterait une dimension importante à l'analyse des données relatives aux gains.

Il faudrait chercher à obtenir des données sur les taux de rémunération. Cet aspect est plutôt difficile à aborder et peut être envisagé de diverses manières. La façon la plus directe consiste à inclure une question sur le taux régulier de rémunération. Une autre possibilité serait de poser une question sur le nombre de semaines travaillées et une autre sur le nombre habituel d'heures de travail par semaine durant la période visée. À cause du problème d'interprétation des chiffres relatifs aux traitements et salaires, il demeurerait difficile de produire un chiffre correspondant au taux de rémunération.

Par ailleurs, étant donné que 40% des employés non agricoles sont syndiqués et que près de la moitié de la main-d'oeuvre est régie par des conventions collectives, il serait bon d'inclure une question sur l'appartenance à un syndicat ou à une association professionnelle.

On pourrait également envisager la possibilité d'inclure une question sur la part de l'employeur correspondant aux congés, aux jours fériés, aux régimes d'assurance-maladie, de retraite, d'avantages sociaux, etc. Bien que ce sujet soit d'importance puisque le coût de ces avantages sociaux représente de 30% à 40% des coûts directs au titre des salaires et traitements, il peut s'avérer difficile d'inclure dans le questionnaire du recensement une question précise.

Enfin, un dernier aspect à prendre en considération est la nécessité de préserver la comparabilité des séries. Les suggestions que nous venons de faire ne risquent pas d'avoir une incidence sur ces dernières car, du fait qu'il s'agit essentiellement d'ajouts ou de subdivisions (dans le cas des données sur l'emploi principal et les emplois secondaires), la comparabilité avec les séries antérieures peut être rétablie.

L'élément le plus important en ce qui a trait aux données sur le marché du travail est la classification des professions et, à un moindre degré, la classification des industries. Pour le recensement de 1991, on utilisera la Classification type des industries de 1980 dont les catégories peuvent être regroupées pour permettre la comparaison avec la classification de 1970 qui a servi aux recensements de 1971 et 1981. La classification des professions est le point le plus délicat. C'est le pivot de toute analyse du marché du travail. Jusqu'à 1981, la classification des professions a subi d'importantes modifications à chaque recensement. La révision la plus conséquente a été celle de 1971, cette version ne comptant que 9 catégories sur 489 parfaitement comparables aux catégories utilisées au recensement de 1961 (Meltz et Stager, 1979). Fort heureusement, les catégories adoptées en 1981 étaient presque entièrement comparables. Qu'en sera-t-il en 1991?

Il est impératif que notre groupe fasse des pressions pour que la comparabilité avec la

classification des professions utilisée en 1971 soit maintenue ou que, du moins, un ajustement des catégories soit possible. Le professeur David Foot et moi-même venons de terminer, pour le compte d'Emploi et Immigration, une étude des déterminants économiques qui sont à l'origine des changements observés dans la composition professionnelle de l'emploi. Nous avons également évalué la précision des classifications des professions utilisées antérieurement. Notre travail a consisté principalement à déterminer quels changements observés dans la composition professionnelle étaient réels et lesquels étaient dus à une révision de la classification. Les principales projections pour le Canada faites dans les années 1960 et au début des années 1970 (Meltz et Penz, 1968; Ahamad, 1969) étaient fondées sur des classifications qui sont devenues périmées avec les recensements subséquents.

L'analyse des ressources humaines et les sujets connexes, comme l'opposition temps partiel et plein temps, la situation vis-à-vis de l'activité dans le cas des personnes âgées, divers aspects de l'immigration, l'égalité d'accès à l'emploi, la parité de salaire à travail égal, etc., supposent une grande cohérence de la classification des professions. Si j'étais limité à une seule proposition, je la formulerais comme suit: de grâce, faites en sorte que la classification des professions de 1991 soit comparable à celle des bases de données de 1981 et 1971.

BIBLIOGRAPHIE

Ahamad, Bill, 1969. **Projection des besoins en main-d'oeuvre par profession en 1975: le Canada et ses régions.** Ottawa: Imprimeur de la Reine.

Meltz, Noah M. et Penz, G. Peter, 1968. **Canada's Manpower Requirements in 1970**, ministère de la Main-d'oeuvre et de l'Immigration. Ottawa: Imprimeur de la Reine.

Meltz, Noah M. et Stager, David A.A., 1979. **Structure de la rémunération selon les professions au Canada, 1931-1975**, Commission de lutte contre l'inflation du Canada. Hull (Québec): ministère des Approvisionnements et Services.

SÉANCE: RECENSEMENT DE L'AGRICULTURE

Président: Terry Gigantes
Direction de la statistique sur les ressources, la technologie et les services
Statistique Canada

Mercredi, le 9 octobre 1985



DÉNOMBREMENT POSTAL DU RECENSEMENT DE L'AGRICULTURE AUX ÉTATS-UNIS

CYNTHIA Z. F. CLARK

U.S. BUREAU OF THE CENSUS

PRÉSENTÉ PAR CHARLES P. PAUTLER

AGRICULTURE DIVISION
U.S. BUREAU OF THE CENSUS

1. Historique du recensement de l'agriculture

Le recensement de l'agriculture aux États-Unis sert à brosser un tableau statistique détaillé d'un secteur vital de l'économie du pays. En règle générale, le recensement a lieu tous les cinq ans et fournit des données sur les terres faisant partie des exploitations agricoles, la production agricole, les ventes de produits agricoles et les caractéristiques des exploitants agricoles, ainsi que d'autres types de données relatives à ce secteur. Ces renseignements sont utilisés par les exploitants agricoles et les responsables des organismes gouvernementaux et privés dans le cadre de la prise de décisions; ils servent aussi de données repères pour les enquêtes et de données de base pour la recherche de nouveaux marchés agricoles. Le présent exposé décrit les procédures adoptées pour le recensement de l'agriculture aux États-Unis que Statistique Canada pourrait appliquer au recensement canadien de 1991.

Depuis 1840, date du premier recensement de l'agriculture aux États-Unis qui avait été conçu comme faisant partie intégrante du recensement décennal, vingt-deux recensements de l'agriculture ont été tenus dans ce pays. De 1840 à 1950, le recensement de l'agriculture a fait partie du recensement décennal. En 1925, 1935 et 1945, un recensement de l'agriculture a été tenu séparément au milieu de la décennie. De 1954 à 1974, le recensement de l'agriculture a été effectué chaque année se terminant par 4 ou 9. En 1976, le Congrès a autorisé la tenue du recensement de l'agriculture en 1978 et en 1982 et tous les cinq ans par la suite de façon que celui-ci coïncide avec les recensements relatifs à l'économie, couvrant les secteurs de la fabrication, des entreprises, des administrations publiques, du transport et de la construction. Le fait d'avoir modifié la période de référence a accru la comparabilité des données ainsi que l'efficience du dépouillement des données de ces recensements.

Le recensement de l'agriculture est imposé par la loi , comme le stipule l'article 13 du Code des États-Unis qui régit les activités du Census Bureau, organisme qui relève du U.S. Department of Commerce. Pour assurer la confidentialité des données, il est interdit d'utiliser ces dernières à des fins autres que statistiques ou de diffuser des données qui permettraient d'identifier des particuliers et l'accès aux rapports du recensement est limité aux gestionnaires et aux employés assermentés. La loi précise l'unité qui doit être utilisée pour la collecte des données relatives aux exploitations agricoles. Depuis 1850, date à laquelle des critères minimaux ont été établis pour définir l'exploitation agricole aux fins du recensement, la définition a été modifiée neuf fois. Selon la définition en usage actuellement, utilisée pour la première fois dans les rapports finals du recensement de 1974, une exploitation agricole est une exploitation qui a vendu, ou qui aurait pu vendre, pour $1,000 ou plus de produits agricoles durant l'année visée par le recensement. Une exploitation dont le chiffre de ventes n'atteint pas la limite fixée peut néanmoins être considérée comme une exploitation agricole si le chiffre des ventes potentielles, établi à partir du dénombrement des cultures et du cheptel et de la production correspondante, répond au critère.

Le recensement de l'agriculture est la principale source de statistiques sur ce secteur de l'économie du pays et la seule source de données sur l'agriculture qui soient cohérentes et comparables aux niveaux des comtés, des états et de l'ensemble du pays. Les données du recensement sont utiles non seulement aux exploitants agricoles mais aussi à l'ensemble du secteur agrocommercial de notre économie. Ces données, de même que les estimations courantes établies à partir d'échantillons et dérivées des données repères tirées du recensement, ou basées sur ces dernières, sont largement utilisées à des fins de planification par les manufacturiers dont les produits visent la clientèle agricole et par les entreprises qui assurent le transport, le conditionnement ou la distribution des produits agricoles aux consommateurs. Les données du recensement sont également utilisées par le Congrès lorsqu'il s'agit d'élaborer des programmes agricoles et d'en analyser les résultats. Bon nombre de programmes

à portée nationale ou à l'échelle des états visant le secteur agricole sont élaborés, ou encore les fonds nécessaires sont alloués, sur la base des résultats du recensement; c'est le cas, notamment, des fonds versés pour l'extension des services, pour la recherche et pour les projets de conservation du sol. Chaque exploitant agricole peut comparer ses propres activités agricoles aux chiffres totaux ou moyens établis pour son comté.

Les données recueillies concernant les activités agricoles englobent les éléments suivants: superficie, utilisation des terres, irrigation; cultures, c'est-à-dire grandes cultures, terres à foin et légumes; bétail, volaille et produits d'origine animale; ventes de produits agricoles; caractéristiques des exploitants agricoles; utilisation des engrais, pesticides et autres dépenses au titre de l'énergie; installations et machines agricoles, et valeur marchande des terres et bâtiments. En plus de ces données, le recensement de l'agriculture fournit une liste des exploitants visés par le recensement qui ont répondu au questionnaire, liste qui est utilisée comme base de sondage pour recueillir des données plus spécialisées sur certains secteurs de l'économie agricole. Antérieurement, ces données ont été recueillies au moyen des enquêtes suivantes: 1979 Farm Finance Survey, 1979 Farm Energy Survey, Surveys of Farm and Ranch Irrigators de 1979 et 1984.

Au recensement de l'agriculture de 1982, on a dénombré 2.2 millions d'exploitations agricoles aux États-Unis. Ces exploitations totalisaient 987 millions d'acres de terre et 132 milliards de dollars au titre de la vente de produits agricoles[1]. D'après les résultats du programme d'évaluation de la couverture, le recensement de 1982 a permis de dénombrer environ 91% des exploitations agricoles à travers tout le pays[2]. La couverture a été bien meilleure dans le cas des exploitations dont le chiffre des ventes était de $2,500 ou plus que dans celui des exploitations dont les ventes étaient inférieures à $2,500.

Les résultats du recensement de l'agriculture sont mis à la disposition des utilisateurs de diverses façons. Dans le cas du recensement de l'agriculture de 1982, les résultats provisoires et finals ont été fournis pour chaque comté, chaque état et pour l'ensemble du pays. Les données finales ont été rendues disponibles pour les régions éloignées, c'est-à-dire Porto-Rico, Guam et les Îles Vierges. Les rapports préliminaires et finals ont été diffusés sous forme de publications et de bandes magnétiques. Les tableaux de données provisoires ont également été fournis sur disquettes pour microordinateur contenant des tableaux détaillés pour chaque comté, chaque état et pour l'ensemble du pays.

Quatre autres publications des données de 1982 sont disponibles sur papier. La publication "Ranking Counties and States" précise l'ordre de grandeur pour certaines données. La publication "Graphic Summary" donne une présentation illustrée du secteur agricole américain au moyen de cartes utilisant les points et les couleurs. La publication "Coverage Evaluation" donne des estimations de l'intégralité de la couverture du recensement pour les États-Unis et pour quatre régions géographiques. Enfin, la publication "Procedural History" est un sommaire détaillé des activités relatives à la planification, à la préparation, à la collecte, au dépouillement et à la diffusion des données.

2. Dénombrement postal pour le recensement de l'agriculture

2.1 Historique du dénombrement

Le recensement de l'agriculture de 1969 a été le premier recensement national de l'agriculture utilisant la méthode d'envoi et de retour du questionnaire par la poste. Tous les recensements antérieurs avaient été effectués par interview sur place dans toutes les régions rurales. Avant 1950, le recenseur devait se baser sur la définition de l'exploitation agricole pour faire remplir un questionnaire aux exploitants dont l'exploitation répondait à la définition. À partir de 1950, le recenseur a eu pour instruction de faire remplir un questionnaire pour toute exploitation où l'on s'adonnait à certains types d'activités agricoles. C'est à l'occasion du dépouillement des questionnaires au bureau central que l'on déterminait s'il s'agissait ou non d'une exploitation agricole. Cette façon de faire devait permettre d'améliorer la couverture des activités qui étaient à l'origine d'une partie importante du sous-dénombrement des exploitations agricoles.

En 1954, un registre a été utilisé dans chaque district de dénombrement en vue d'améliorer la couverture. Le recenseur devait consigner l'emplacement et les données d'identification

[1] **1982 Census of Agriculture**, U.S. Summary and State Data, vol. 1, partie 51.

[2] **1982 Census of Agriculture**, volume 2, partie 2, Coverage Evaluation.

de chaque logement ou de chaque endroit où des activités agricoles étaient effectuées, sans qu'il y ait de logement sur place, dans la mesure où l'endroit en question était situé en partie ou en totalité dans le district de dénombrement. Le recenseur devait, en outre, dessiner les limites de chaque secteur agricole et non agricole sur une formule illustrant le canton (township).

Étant donné la difficulté que l'on avait à trouver des recenseurs pour dénombrer les régions rurales et le coût du dénombrement par interview sur place, on a étudié la possibilité de recourir à d'autres méthodes de collecte des données. À la suite du recensement de 1964, une étude spéciale a été effectuée pour évaluer la faisabilité de la collecte de données par la poste basée sur une liste d'adresses établie à partir des déclarations de revenus soumises à l'administration fédérale. Un échantillon des exploitants agricoles dénombrés lors du recensement de 1964 a été apparié aux déclarations de revenus produites en 1963 et en 1964. Pour chaque exploitant agricole faisant partie de l'échantillon, on a déterminé si une formule 1040F de déclaration de revenus avait été reçue en 1963 ou en 1964. L'étude a révélé que, pour l'ensemble du pays, environ 96% des exploitations dont la valeur totale de produits agricoles (VTP) était de $2,500 ou plus avaient des chances de figurer sur une liste d'adresses établie à partir des données de l'Internal Revenue Service (IRS). Dans le cas des exploitations dont la VTP n'atteignait pas $2,500, environ 70% d'entre elles avaient des chances de figurer sur la liste de l'IRS. Compte tenu des résultats de cette étude, il a été décidé d'adopter la méthode d'envoi et de retour du questionnaire par la poste pour le recensement de 1969 en utilisant les listes d'exploitants agricoles établies à partir des déclarations de revenus et d'autres listes agricoles.

2.2 Élaboration de la liste d'adresses

Dans le cadre de la mise sur pied des procédures de collecte des données au moyen de l'envoi et du retour du questionnaire par la poste, un processus d'élaboration de la liste d'adresses a été prévu[3]. Étant donné que l'intégralité des données du recensement dépend dans une grande mesure de l'intégralité de la liste d'adresses, il s'agit d'une des étapes les plus importantes de la tâche relative à la tenue du recensement de l'agriculture. L'objectif du processus d'élaboration de la liste d'adresses est d'établir une liste exhaustive qui évite les doublets et qui ne répertorie que les exploitations agricoles. La liste d'adresses établie pour le recensement de l'agriculture est le premier exemple d'une telle liste établie à partir de multiples registres administratifs. La disponibilité de registres administratifs et l'accessibilité de ces derniers sont des conditions essentielles à la collecte des données sur l'agriculture par voie postale.

Les noms et adresses de personnes et organismes ayant un lien avec l'agriculture sont obtenus de plusieurs sources principales. En 1969, ces sources incluaient le fichier de l'IRS sur les formules 1040F établies pour des exploitations agricoles, le fichier des formules 1065 établies pour des exploitations agricoles en association, le fichier des formules 1120S établies pour de petites exploitations agricoles en corporation, le fichier des formules 943 établies pour des employés agricoles, le fichier de l'Agriculture Stabilization and Conservation Service (ASCS), la liste de l'United States Department of Agriculture's Statistical Reporting Service (SRS) dans les états du Nord-est et du Sud, et dans certains états du Nord de la région centrale, et la liste des exploitations de grande taille, établie à partir des résultats des recensements antérieurs, dûment mise à jour. Depuis 1969, on a ajouté à ces sources les fichiers des exploitants agricoles établis à partir des résultats du recensement de l'agriculture précédent, le fichier des non-répondants du recensement précédent, la liste du SRS pour tous les états répertoriés et des listes spéciales obtenues de diverses sources énumérant les exploitations de grande taille ou les exploitations spécialisées. Toutefois, toutes les exploitations répertoriées sur ces diverses listes ne répondent pas nécessairement à la définition établie pour le recensement.

La qualité des renseignements relatifs aux identificateurs utilisés sur les listes obtenues des diverses sources extérieures du Census Bureau varie selon la source. La plupart de

---

3 Dea, Jane Y., Gaulden, Tommy W. et Prochaska, D. Dean, "Record Linkage for the 1982 Census of Agriculture Mail List Development Using Multiple Sources", 1984 **Proceedings of the Section on Survey Research Methods**, American Statistical Association.

ces listes porte un code ou un chiffre indiquant la taille de l'exploitation et son type. Il y a énormément de doublets entre et à l'intérieur de ces fichiers. Les listes fournies par ces sources peuvent contenir des variations du même nom, qu'il s'agisse du surnom, des initiales, du deuxième prénom de l'exploitant ou du nom de l'exploitation. Les exploitants agricoles peuvent avoir différentes adresses, pour l'exploitation proprement dite et pour leur lieu de résidence ou par suite d'un déménagement. Les renseignements qui ne concernent pas l'exploitation agricole et les enregistrements en double provenant du recensement précédent servent à déterminer ce qui correspond à l'exploitation agricole et ce qui fait double emploi.

L'élaboration de la liste d'adresses pour le recensement comprend deux phases: 1) Farm and Ranch Identification Survey (15.8 millions de dossiers source en 1982) et 2) le recensement postal (3.2 millions de dossiers source supplémentaires en 1982). Chaque phase comprend cinq grandes opérations: 1) format et normalisation; 2) couplage du numéro d'identification de l'employeur (EIN) et du numéro de sécurité sociale (SSN); 3) codage géographique et contrôle du code postal; 4) couplage alphabétique des noms; et 5) revue manuelle de tous les ensembles d'enregistrements qui n'ont pas été identifiés comme étant ou non des doublets. Le couplage alphabétique des noms est basé sur la théorie du couplage des enregistrements élaborée par Fellegi et Sunter de Statistique Canada.

En 1982, la première phase du couplage des enregistrements a abouti à la création d'un fichier d'environ 7.3 millions d'enregistrements. Ces enregistrements ont été classés en trois catégories, soit comme "exploitation agricole probable" (1.9 million d'enregistrements), soit comme "exploitation agricole discutable" (3.1 millions d'enregistrements) ou soit comme "exploitation agricole improbable" (2.3 millions d'enregistrements). Les enregistrements de la catégorie "exploitation agricole improbable" ont été éliminés de la liste d'adresses; ceux de la catégorie "exploitation agricole discutable" ont été sélectionnés pour Farm and Ranch Identification Survey. L'objectif de cette enquête est d'identifier les exploitants non agricoles et d'ajouter les noms des nouveaux locataires ou des successeurs. Les enregistrements qui ont été identifiés comme des exploitations agricoles visées par le recensement à l'occasion de cette enquête, ceux qui proviennent de sources antérieurement non disponibles et les enregistrements de la catégorie "Exploitation agricole probable" ont été regroupés lors de la deuxième phase du couplage des enregistrements pour constituer la liste d'adresses finale. En 1982, ce processus à deux phases a abouti à l'établissement d'une liste d'adresses correspondant à 3.6 millions d'enregistrements à partir d'un total de 19 millions d'enregistrements source.

Pour le couplage des données de 1982, l'utilisation d'enregistrements se rapportant à des exploitations non agricoles tirés du recensement précédent a permis de réduire la longueur de la liste d'adresses finale et d'éliminer bon nombre d'exploitations non agricoles. La liste finale de 1982 compte environ 20% moins d'enregistrements que celle de 1978 qui en totalisait 4.4 millions. Une étude de la comparabilité des données de recensement a révélé que la couverture des listes d'adresses finales de 1978 et 1982 était passablement comparable[4]. Le coût du couplage des enregistrements et de l'élaboration de la liste d'adresses de 1982 s'est monté à près de 1.5 million de dollars. Il faut ajouter à cela une dépense supplémentaire de 4 millions de dollars occasionnée par la tenue de Farm and Ranch Identification Survey.

2.3 Limites du dénombrement postal des exploitants agricoles

Bien que la liste d'adresses finale établie pour le recensement soit compilée à partir d'un nombre important de listes source, elle ne couvre pas parfaitement l'univers des exploitations agricoles visées par le recensement. Parmi les types d'exploitations qui ne sont pas toujours répertoriés sur les listes source, citons les exploitations dont l'exploitant n'est que faiblement impliqué dans des activités agricoles et celles qui ne sont pas déclarées comme exploitations agricoles par leur exploitant. Une exploitation peut donc être classée par erreur dans la catégorie "exploitation agricole improbable"

[4] Clark, Cynthia Z.F., "Comparability of Data from the Censuses of Agriculture",1984 **Proceedings of the Section on Survey Research Methods**, American Statistical Association.

et éliminée de la liste avant que l'on procède à Farm and Ranch Identification Survey. Il peut aussi arriver qu'une exploitation soit éliminée de la liste parce que l'exploitant a mal répondu aux questions de sélection posées lors de l'enquête ou parce que la réponse qui a été fournie dans le questionnaire de recensement a été mal interprétée de sorte que l'exploitation a été classée comme exploitation non agricole. Toutes ces erreurs sont à l'origine du "sous-dénombrement" des exploitations agricoles visées par le recensement. On estime à 336,000 le nombre d'exploitations agricoles ainsi oubliées en 1982[5].

Le processus d'élaboration de la liste d'adresses entraîne un autre problème qui doit être pris en considération dans les procédures de recensement. Comme il a été souligné plus haut, plusieurs listes source contiennent des noms ou des adresses en double. Le couplage des enregistrements permet d'éliminer la plupart de ces doublets. Les recensés sont invités à appliquer une procédure spéciale lorsqu'ils reçoivent le questionnaire en double. Toutefois, cela n'empêche pas qu'un certain nombre de cas de doublets persiste. Ces derniers, combinés avec les exploitations non agricoles qui ont été classées par erreur comme des exploitations agricoles lors du dépouillement des données de recensement, sont à l'origine du nombre estimatif d'exploitations agricoles dénombrées en trop, qui était de 114,000 en 1982[6].

Comme il a été souligné précédemment, les évaluations de la couverture indiquent que le dénombrement est particulièrement incomplet dans le cas des exploitations agricoles dont le chiffre des ventes est inférieur à $2,500. Nombre de ces petits exploitants agricoles n'ont aucun lien formel avec des organismes agricoles ou ne s'identifient aucunement comme des exploitants agricoles. De ce fait, leur nom ne figure pas sur les listes source qui servent à constituer la liste d'adresses pour le recensement. Il faut donc adopter une autre méthode pour assurer une meilleure couverture de cette portion de l'univers des exploitations agricoles.

En 1978, le recensement de l'agriculture incluait un échantillon aréolaire de secteurs dans le dénombrement postal[7]. Les données tirées de l'échantillon aréolaire ont été ajoutées à celles du recensement dans les répartitions aux niveaux des états et de l'ensemble du pays.

Cette façon de faire a permis d'accroître sensiblement la couverture du recensement, notamment en ce qui a trait aux exploitations dont le chiffre des ventes est inférieur à $2,500. En 1978, le pourcentage des petites exploitations oubliées était de 3.2 alors qu'il était de 28.5 en 1982 (tableau 4). Cependant, cette méthode n'a pu être reprise en 1982 ni en 1987 à cause des restrictions budgétaires.

Quelle que soit la méthode de collecte des données, le taux de réponse influe sur la qualité des données finales. Il est particulièrement difficile d'obtenir que tous les recensés remplissent et retournent leur questionnaire dans le cas du dénombrement postal du fait que la liste d'adresses contient des exploitations agricoles mais aussi des exploitations non agricoles. Les procédures générales de recensement qui sont décrites à la section 3.2 visent à favoriser une bonne participation des recensés. Toutefois, certains types d'exploitations agricoles ont été repérés comme nécessitant l'adoption de procédures spéciales en vue d'obtenir des réponses fournies directement par les intéressés (plutôt que des réponses imputées) ainsi que l'intervention d'un analyste agricole. Ces exploitations agricoles de grande taille ou très spécialisées, qui doivent absolument être dénombrées individuellement, sont indiquées par une mention particulière sur l'étiquette d'adresse afin de faciliter l'utilisation des procédures correspondantes. Elles incluent les exploitations à établissements multiples, les exploitations non conventionnelles et celles pour lesquelles le chiffre estimatif des ventes dépasse le seuil établi (habituellement $100,000).

La plupart des données publiées dans le cadre du recensement de l'agriculture ont été recueillies auprès de l'ensemble des exploitants répertoriés sur la liste d'adresses. Toutefois, certaines données sont fournies par un échantillon de recensés tiré de la liste d'adresses. Il s'agit notamment des données sur les engrais et les insecticides, sur les installations et machines agricoles, sur les dépenses au titre des intérêts ou de l'énergie,

5 **1982 Census of Agriculture**, volume 2, partie 2, Coverage Evaluation.

6 ibid.

7 **1978 Census of Agriculture**, volume 5, partie 4, Procedural History.

sur la production et sur la valeur marchande des terres et bâtiments. En raison de l'incidence des réponses fournies pour les exploitations de grande taille sur les chiffres estimatifs correspondant à ces catégories de données, l'échantillon a été tiré de façon à inclure en toute certitude certaines exploitations figurant sur la liste d'adresses qui sont susceptibles de répondre aux critères fixés en matière de taille (superficie ou chiffre des ventes) ou de limites géographiques. En ce qui a trait à la taille, les critères fixés varient selon les états, cela peut aller de 1,000 à 5,000 acres et de plus de $40,000 à plus de $200,000 de chiffre des ventes. Toutes les exploitations des comtés où l'on avait dénombré moins de 100 exploitations agricoles au recensement précédent ont été incluses dans l'échantillon. Il y a chevauchement du groupe des exploitations qui doivent être dénombrées individuellement et de celui des exploitations qui sont incluses avec certitude dans l'échantillon du fait que les exploitations du premier groupe, qui ont été sélectionnées parce qu'elles répondaient au critère relatif au chiffre des ventes, sont nécessairement dans la strate correspondant au second groupe.

3. Procédures utilisées au dénombrement postal

3.1 Publicité et programme de sensibilisation du public

Le programme du recensement de l'agriculture utilise diverses approches publicitaires en vue de rejoindre tous les niveaux du secteur agricole. Avant le recensement, des affiches et des trousses d'information sont préparées à l'intention de ceux qui acceptent d'aider à promouvoir le recensement. Le Bureau demande aux responsables des établissements commerciaux et bancaires locaux de placer les affiches dans des emplacements stratégiques. Il envoie également des membres de son personnel assister à des expositions commerciales et agricoles afin de faire connaître le recensement aux exploitants agricoles et de renseigner les utilisateurs au sujet des données du recensement. À ces occasions, des brochures et un feuillet sont distribués pour souligner l'importance du questionnaire intitulé "Fill it out - Mail it back".

Le Bureau sollicite la collaboration d'organismes agricoles des comtés tels que le County Extension Service, l'Agriculture Stabilization and Conservation Service et le Soil Conservation Service pour la diffusion de l'information et la promotion du recensement aux réunions d'exploitants agricoles, mais aussi pour aider les recensés à remplir le questionnaire de recensement. Un guide est également rédigé à l'intention des responsables de ces organismes leur donnant les instructions relatives à la manière de remplir le questionnaire de recensement. En outre, des documents sont rédigés spécifiquement pour les étudiants en agriculture en vue de leur insertion dans les programmes de formation agricole. Des guides et des plans de leçon sont prévus pour exposer l'importance du recensement de l'agriculture et inviter les étudiants à aider leurs parents à remplir le questionnaire de recensement. Pour les recensements antérieurs, le Bureau a bénéficié de la publicité favorable faite par les membres du Congrès, du Sénat et des assemblées législatives des divers états, et les politiciens locaux ont donné un support verbal au recensement de l'agriculture.

Le Bureau sollicite également la collaboration des responsables de stations de radio, de revues agricoles et de magazines. Ces médias atteignent la plupart des occupants des ranches et des exploitations agricoles. Les annonceurs d'émissions de radio et de télévision destinées au monde agricole contribuent à promouvoir le recensement en informant les exploitants agricoles et les exploitants de ranches au sujet de cet événement et des données de recensement, à l'occasion des communiqués et autres types de nouvelles agricoles. Le Bureau fournit des renseignements qui feront la matière de courts bulletins de nouvelles et d'encarts publicitaires destinés à être insérés dans les revues et journaux, à portée locale ou nationale, visant le monde agricole. Il demande aux rédacteurs en chef des magazines de publier des articles traitant du recensement et, dans la mesure du possible, de mettre une photographie en couverture en guise d'introduction. Il fournit également de l'information et de la publicité aux associations professionnelles pour leurs publications. Ces dernières annoncent les expositions agricoles et autres événements d'intérêt spécifique et sont lues par divers secteurs de l'univers agricole.

Les différentes méthodes que nous venons d'énumérer s'avèrent particulièrement utiles lorsqu'il n'y a pas de contact direct avec les recensés par l'intermédiaire d'un recenseur. Le premier envoi postal contient, en plus du questionnaire de recensement, une brochure et un feuillet expliquant la raison d'être du recensement et son utilité pour les recensés.

Les envois subséquents, faits à l'occasion du suivi, contiennent des renseignements additionnels.

3.2 Collecte des données

Pour les recensements de 1969, 1974, 1978 et 1982, les envois postaux étaient composés d'une lettre, du questionnaire et d'une fiche de rappel. Les procédures adoptées pour ces divers recensements étant similaires, nous en donnerons une présentation détaillée pour le recensement de 1982 seulement[8]. Les questionnaires de recensement ont été expédiés à la fin de décembre 1982 et devaient être retournés à la mi-février. Il était prévu d'envoyer jusqu'à six fiches de rappel, à un intervalle de trois semaines environ. Le premier suivi consistait à envoyer, après la date limite de la mi-février, une fiche de rappel à tous les recensés qui n'avaient pas retourné leur questionnaire. Les deuxième et sixième envois effectués dans le cadre du suivi contenaient un questionnaire de recensement ainsi que les instructions pertinentes. Les autres envois contenaient une lettre invitant le recensé à remplir son questionnaire et à le retourner, faisant état des diverses utilisations des données du recensement et rappelant aux recensés que la Loi les oblige à participer au recensement. Le tableau 1 donne la date de chaque envoi, le contenu et le nombre d'envois effectués ainsi que le taux de réponse au moment de l'envoi.

À l'occasion du deuxième envoi effectué dans le cadre du suivi, un fichier a été constitué, en vue du suivi téléphonique ultérieur, avec les

TABLEAU 1. Contact postal pour le recensement de l'agriculture de 1982

Contact postal	Date d'expédition	Contenu de l'envoi	Nombre d'envois	Taux de réponse pour le dénombrement postal (%)	Augmentation du taux de résponse (%)
Premier envoi	Déc. 1982	Lettre, questionnaire et copie du fichier, instructions, brochure "Your Farm or Ranch Counts"	3,600,000		
	1/28/1983			31.4	31.4
Premier suivi (date limite 2/15/1983)	2/22/1983	Fiche de rappel	1,900,000	48.4	17.0
Deuxième suivi	3/15/1983	Lettre, questionnaire, instructions	1,600,000	57.3	8.9
Troisième suivi	4/13/1983	Lettre	1,071,000	70.1	12.8
Quatrième suivi	5/12/1983	Lettre	890,000	75.6	5.5
Cinquième suivi	5/25/1983	Lettre	790,000	78.2	2.6
Sixième suivi	6/21/1983	Lettre, questionnaire, instructions	708,000	80.3	2.1
	7/15/1983			83.0	2.7
	9/09/1983			85.2	2.2
	Date finale			85.4	.2

Source: U.S. Bureau of the Census, données finales tirées des tableaux relatifs au dénombrement postal fournies par l'Agriculture Division.

[8] Ruggles, Donna R., Dea, Jane Y., Kwok, Flora K., et Carman, Cindy A., "Evaluation of the Effectiveness of Data Collection Procedures for the 1982 Census of Agriculture", 1984 **Proceedings of the Section on Survey Research Methods**, American Statistical Association.

noms et adresses des recensés faisant partie du groupe des exploitants agricoles qui devaient être dénombrés individuellement et qui n'avaient pas retourné leur questionnaire rempli. Ce fichier comprenait 417,000 exploitations agricoles dont le chiffre des ventes potentielles pour 1982 était de $100,000 ou plus. Les noms contenus dans ce fichier ont été conservés dans le fichier du suivi postal jusqu'à ce que l'exploitant ait communiqué ses réponses au questionnaire par téléphone ou par la poste. Pour les exploitants devant être dénombrés individuellement, le suivi téléphonique a débuté en mai. Lorsque aucune réponse n'était fournie par téléphone ni par la poste dans le cas des non-répondants qui appartenaient à ce groupe, le Bureau essayait d'obtenir les données requises d'une autre source d'information. Les 2,700 bureaux locaux de l'Agriculture Stabilization and Conservation Service (ASCS) du U.S. Department of Agriculture étaient la principale source de renseignements consultée. Les questionnaires ainsi remplis étaient ensuite vérifiés par un analyste.

Bien que le taux de réponse global au recensement ait atteint 65% en avril 1983, dans un certain nombre de comtés le taux était de beaucoup inférieur. Pour favoriser la participation dans ces régions, il a été décidé de procéder à un suivi supplémentaire auprès des non-répondants dans certains états ou comtés où le taux de réponse était inférieur au taux enregistré pour l'ensemble du pays. Une lettre de suivi spéciale a été envoyée par la poste à ce groupe de 309,000 non-répondants leur offrant de l'aide pour remplir le questionnaire de recensement. En mai 1983, un suivi téléphonique supplémentaire a été effectué auprès d'un échantillon de non-répondants de 252 comtés où le taux de réponse était inférieur à 75%. Le service téléphonique de Jeffersonville, Indiana, prévu au départ pour répondre essentiellement aux appels reçus des recensés, a été chargé par la suite du suivi téléphonique. Dans la mesure du possible, les préposés ont essayé d'avoir le numéro de téléphone des non-répondants et d'effectuer des interviews téléphoniques. Lorsqu'il était impossible d'obtenir un questionnaire rempli pour des exploitations devant être dénombrées individuellement, les préposés au service téléphonique essayaient d'obtenir les données requises d'autres sources de renseignements. Ils communiquaient également avec les exploitants du même groupe qui n'avaient que partiellement rempli leur questionnaire.

3.3 Contrôle, imputation et revue analytique

Les questionnaires retournés par la poste étaient vérifiés au centre de dépouillement.

Lorsque les codes ne pouvaient être lus par la machine, les données à vérifier étaient introduites par clavier. Lorsque le questionnaire était en blanc ou lorsque des documents y étaient joints, il était transmis à la section chargée de la révision manuelle pour que le cas soit résolu. Les questionnaires étaient ensuite mis en lots, par état, en vue de l'introduction des données par clavier. Pour la première fois au recensement de 1982, aucune prévérification manuelle n'a été effectuée avant l'introduction par clavier. Les préposés avaient pour instruction d'introduire les données, au moyen de codes appropriés, et d'indiquer toute particularité, telles que l'utilisation d'unités non admises, les colonnes mal remplies, l'ajout de remarques concernant les données, les réponses entre parenthèses ou les doubles réponses. Les préposés devaient même faire état de toute note ou commentaire écrit, en introduisant un code spécial. Ces particularités de même que le codage des réponses relatives aux "autres" cultures et au bétail étaient vérifiés au moment de la revue effectuée à la suite du contrôle complexe.

Une fois l'introduction par clavier terminée, les fichiers étaient transmis au Washington Office pour le formatage. À cette occasion, les enregistrements introduits étaient traités par un programme qui, dans un premier temps, convertissait les données en unités de mesure normalisées. Puis, les données étaient présentées selon le système binaire et le format des enregistrements était normalisé. Tout numéro au fichier de recensement (CFN) comportait une section de longueur fixe correspondant à l'indicateur et une section de longueur variable pour les valeurs correspondant aux données. Cette dernière section comportait des codes spéciaux pour tout élément de données déclaré, changé ou imputé. Les enregistrements non pertinents étaient identifiés. Tout code relatif à l'état, au comté ou au type de formule non valide était identifié en vue d'être corrigé et réintroduit. Un fichier de correction était alors créé pour les codes incorrects et ensuite apparié au fichier principal de façon à remplacer les données erronées par les valeurs appropriées. Lorsque le nombre de rejets dépassait la limite admissible, l'enregistrement était réintroduit intégralement.

Ensuite, les applications détaillées du contrôle et de l'imputation étaient effectuées à l'occasion du contrôle complexe. Cette opération consistait à valider, à contre-vérifier et à raffiner les données fournies. Les spécifications relatives au contrôle étaient données aux programmeurs par les spécialistes. Les ratios d'introduction étaient alors comparés aux limites admissibles établies à partir des résultats obtenus lors des recensements et enquêtes antérieurs. Les programmes de contrôle arrondissaient les données, substituaient la somme des éléments de données aux totaux déclarés et imputaient des valeurs à des éléments de données à partir de ratios établis.

Le programme informatique utilisé pour le contrôle complexe comportait plus de 10,000 étapes, mais toutes n'étaient pas utilisées dans chaque relevé. Les contrôles étaient effectués séparément pour chaque état. À l'occasion du contrôle de chaque enregistrement, la classification de l'exploitation était déterminée, les données manquantes étaient ajoutées (imputation de valeurs aux éléments de données), la superficie totale des terres en culture était ajustée en fonction des superficies déclarées pour chaque culture, les valeurs déclarées qui sortaient des limites établies étaient rectifiées, les montants déclarés au titre des ventes de produits agricoles étaient vérifiés au moyen des prix moyens en vigueur dans l'état, d'autres vérifications de la cohérence des données étaient effectuées et la superficie, le mode d'occupation, la valeur totale des produits agricoles (VTP), le code CTI, ainsi que la forme juridique étaient déterminés. Enfin, les données de 1982 étaient comparées à celles de 1978. Cette façon de faire remplaçait la procédure adoptée au recensement précédent selon laquelle, avant l'introduction des données, un analyste effectuait une revue de tous les questionnaires des exploitations devant être dénombrées individuellement.

Une fois le contrôle informatique complexe terminé, un imprimé d'ordinateur était produit répertoriant tous les éléments de données qui avaient été rejetés au contrôle, les éléments qui avaient été modifiés, ainsi que les indicateurs qui avaient été mis en place lors du contrôle. Ces imprimés de même que les questionnaires correspondants étaient ensuite passés en revue manuellement à Jeffersonville. Lorsque le cas ne pouvait être réglé manuellement, il était soumis à un analyste des données agricoles. Les corrections étaient alors préparées, puis introduites. Le fichier de correction était ensuite transmis au Washington, D.C. Office pour être apparié au fichier principal. Les fichiers corrigés étaient à nouveau soumis au contrôle pour vérifier si les corrections avaient été effectuées convenablement et si d'autres mesures devaient être prises.

Une fois que le contrôle avait été effectué et les corrections subséquentes apportées, les fichiers dûment corrigés pour chaque état étaient versés dans le fichier détaillé pour l'ensemble du pays, par état. Par la même occasion, les enregistrements en double étaient repérés. Exception faite des cas où les enregistrements étaient identifiés comme correspondant à des opérations différentes effectuées sous le même CFN, seul le premier enregistrement était conservé. Le programme de fusion des fichiers produisait également le nombre d'exploitations d'après la taille, la valeur totale des produits agricoles et le type d'exploitation, ces données étant ensuite utilisées dans le programme d'imputation appliqué aux cas de non-réponse.

3.4 Imputation appliquée aux cas de non-réponse

L'effort concerté, tant au niveau du suivi postal que du suivi téléphonique, en ce qui a trait à la collecte des données du recensement de l'agriculture, a abouti à un taux de réponse de 86%. La poursuite du suivi après la période fixée ne donnerait qu'une augmentation minime du taux de réponse. En vue de la publication de données sur tout l'univers agricole, les données recueillies ont été pondérées pour tenir compte des exploitants agricoles non répondants. Une enquête a été menée auprès des non-répondants pour établir des estimations, par état, de la proportion de non-répondants visés par le recensement de l'agriculture par rapport à l'ensemble des non-répondants afin d'imputer des réponses aux questionnaires des exploitants agricoles non répondants sur la base des données fournies par les autres exploitants.

L'échantillon de non-répondants était un échantillon à un seul degré stratifié systématique avec des taux de sélection variables selon la strate et l'état. À l'intérieur de chaque état, les non-répondants étaient répartis en six strates, déterminées en fonction du volume de courrier, du type de registres administratifs source et des codes spéciaux. Environ 13,000 noms et adresses ont ainsi été sélectionnés à partir de l'univers de non-répondants d'avril. Les taux de sélection variables pour chaque état ont été

établis afin d'estimer le nombre de non-répondants visés par le recensement de l'agriculture avec une erreur d'environ 6%.

Le questionnaire utilisé pour l'enquête auprès des non-répondants avait été conçu de façon à distinguer les adresses des recensés qui exploitaient une ferme visée par le recensement de l'agriculture de celles des autres recensés. Un questionnaire était envoyé aux non-répondants faisant partie de l'échantillon à la fin d'avril et un autre deux semaines plus tard. Ensuite, ces exploitants étaient contactés par téléphone. Étant donné que cette enquête était menée en même temps que le recensement, lorsqu'un exploitant agricole faisant partie de l'échantillon des non-répondants avait rempli le questionnaire de recensement, il était éliminé de l'échantillon.

À partir des résultats de l'enquête auprès des non-répondants, on a estimé le pourcentage d'exploitants agricoles non répondants pour chaque état. Puis, on a utilisé un estimateur synthétique pour calculer le nombre estimatif de non-répondants pour chaque strate de chaque comté de l'état. Enfin, on a sélectionné un échantillon de répondants pour représenter les non-répondants. Toutefois, lorsque le non-répondant faisait partie du groupe des exploitants devant être dénombrés individuellement, les réponses n'étaient pas imputées de cette façon mais, comme il a été expliqué plus haut, elles étaient obtenues à l'occasion du suivi téléphonique, auprès de sources secondaires ou encore à partir des données chronologiques.

Une fois le contrôle et l'imputation entièrement terminés, les données de l'échantillon étaient pondérées, puis compilées. Les analystes des données agricoles passaient alors en revue les totalisations, faisaient une critique détaillée des données peu sûres et proposaient des mesures correctives. Les données étaient ensuite revues manuellement à Jeffersonville afin d'éliminer les doublets et de corriger les erreurs commises au moment de l'introduction par clavier, de la déclaration ou du dépouillement. Par la même occasion, on tentait d'obtenir les renseignements manquants auprès des recensés et on corrigeait les enregistrements erronés qui étaient ensuite soumis à un analyste. Des rapports préliminaires étaient alors élaborés et revus, les corrections requises étant apportées au fichier de données aussi souvent que cela était nécessaire pour assurer l'exactitude des enregistrements. Les données prêtes pour la diffusion étaient soumises à une dernière analyse avant que les tableaux finals soient élaborés.

4. Évaluation du dénombrement postal

4.1 Qualité des données de recensement publiées

Il est particulièrement complexe de publier des données de qualité à partir du recensement de l'agriculture du fait que la liste d'adresses utilisée contient un nombre important d'exploitations (près de 1.4 million en 1982) qui ne correspondent pas à la définition d'une exploitation agricole. Au moment de l'élaboration de la liste d'adresses, un certain nombre d'exploitations dont le statut est inconnu sont néanmoins retenues de façon à assurer une couverture plus adéquate de l'univers visé. En conséquence, les procédures de collecte des données doivent être conçues en fonction des exploitations agricoles "discutables" aussi bien que des exploitations agricoles "véritables". Le questionnaire utilisé doit s'appliquer aux deux types de cas si l'on veut obtenir des réponses et si l'on veut que celles-ci puissent être classifiées.

La qualité des statistiques tirées des questionnaires de recensement dépend de plusieurs facteurs. Citons notamment: 1) l'intégralité et l'exactitude de la liste d'adresses des exploitants agricoles; 2) l'efficacité avec laquelle les procédures de collecte des données adoptées permettent d'obtenir des questionnaires remplis des exploitants inscrits sur la liste; 3) la clarté du questionnaire et des instructions pour remplir celui-ci, du fait que cela influe sur la précision des renseignements fournis par les répondants; 4) la précision avec laquelle les exploitations sont classées comme exploitations agricoles ou non agricoles au moment du dépouillement des données; 5) l'intégralité des questionnaires remplis par les répondants; 6) la portée du contrôle et la mesure dans laquelle il a fallu faire appel à des sources secondaires pour imputer une valeur aux éléments de réponse manquants; 7) l'efficacité des procédures de couplage des enregistrements utilisées pour identifier les doublets; et 8) la fiabilité des méthodes adoptées pour estimer les données dans les cas de non-réponse.

Tous ces facteurs, à l'exception du premier, sont sujets à un changement possible de la méthodologie pour le recensement de l'agriculture de 1991, au Canada. Nous exposerons,

dans la présente section, les résultats des évaluations du recensement de l'agriculture de 1982 et des recensements antérieurs portant sur les facteurs 1, 2, 4, 7 et 8. Ces évaluations avaient trait notamment au taux de réponse et aux procédures de suivi (facteurs 2 et 8), à l'exactitude de la catégorisation des réponses (facteur 4), à l'imputation des valeurs dans les cas de non-réponse (facteur 8) et à la couverture du recensement (facteurs 1, 4 et 7).

Par ailleurs, les recherches qui doivent être menées à l'occasion de l'évaluation du recensement de l'agriculture de 1987 fourniront plus de détails concernant les facteurs 2, 3 et 5, ainsi que nous le verrons à la section 5, intitulée "Mesures à prendre en vue du prochain recensement". Des recherches seront faites dans les prochaines années afin d'avoir plus de détails relativement au contrôle et à l'imputation de valeurs aux éléments de réponse manquants (facteur 6) et à l'imputation de valeurs dans les cas de non-réponse (facteur 8). Le programme d'évaluation de la couverture du recensement de l'agriculture de 1987 fournira des détails additionnels concernant les facteurs 1, 4 et 7 puisqu'il analysera l'incidence des modifications apportées aux méthodes, par rapport aux recensements antérieurs, sur l'élaboration de la liste d'adresses, le couplage des enregistrements et les procédures de classification.

4.2 Taux de réponse comme mesure de l'efficacité de la qualité des données

Le taux de réponse obtenu à une enquête est une mesure normalisée de l'efficacité de la collecte de données complètes auprès de l'univers visé par l'enquête. L'examen, sous divers angles, des réponses fournies aux différents recensements donne un certain nombre d'indications de l'efficacité des méthodes de collecte des données utilisées dans le cadre des divers recensements de l'agriculture. Les taux de réponse publiés correspondent au quotient du total des questionnaires reçus (y compris les questionnaires transmis par les services postaux – Post Master Returns ou PMR) divisé par le nombre total d'exploitations répertoriées sur la liste d'adresses. D'après ce mode de calcul, le taux de réponse de 88% enregistré au recensement de 1978 est largement supérieur au taux de 85.4% établi pour le recensement de 1982 et sensiblement supérieur à celui du recensement de 1974 qui était de 87.4% (tableau 2). Étant donné qu'il y a eu proportionnellement plus de PMR en 1978 qu'en 1974 et en 1982, le mode de calcul du taux de réponse contribue à surestimer dans une certaine mesure l'efficacité de la collecte des données en 1978 par rapport à 1974 et 1982. Si l'on élimine ces PMR du total des questionnaires reçus et du nombre total d'exploitations répertoriées sur la liste d'adresses, on obtient des taux de réponse de l'ordre de 85.1% pour 1982, 87.3% pour 1978 et 87.1% pour 1974. D'après ce dernier mode de calcul, le taux de réponse pour 1982 était de 2.1% inférieur à celui de 1978[9].

L'ampleur de l'univers couvert par la définition d'une exploitation agricole complique la collecte des données recherchées. L'envoi initial du questionnaire et de la lettre de rappel de même que les suivis postaux subséquents ne suffisent pas nécessairement à convaincre leurs destinataires de la nécessité de remplir le questionnaire, que ceux-ci jugent que leurs activités sont d'ordre agricole ou non. En conséquence, un certain nombre d'exploitations agricoles et non agricoles peuvent n'être pas déclarées. Afin de déterminer s'il s'agit d'exploitations agricoles ou non agricoles, il serait probablement plus efficace de procéder par interview téléphonique pour obtenir les renseignements nécessaires des non-répondants qui n'ont pas jugé que leurs activités étaient d'ordre agricole. Cette procédure a été utilisée en 1982 pour obtenir les renseignements auprès des exploitants dont on pensait que l'exploitation était de grande taille ou spécialisée.

Les données finales du recensement de l'agriculture de 1982 sont tirées des 3.1 millions de questionnaires obtenus au moyen d'une liste de 3.6 millions de noms et adresses. Sur l'ensemble de ces répondants, 67.4% étaient des exploitants agricoles. Plusieurs facteurs liés aux procédures adoptées peuvent avoir influé sur le taux de réponse obtenu pour le recensement de l'agriculture, ainsi qu'en témoignent les différences entre les taux de réponse selon le type et la fréquence du suivi. Un autre facteur qui pourrait influer sur le taux de réponse est la longueur du questionnaire. L'évolution des taux de réponse d'un recensement à un autre selon la catégorie de taille d'après la liste d'adresses (mesure de la taille dérivée des indicateurs présents dans les enregistrements obtenus à l'aide de la liste d'adresses) a été analysée en

[9] Clark, Cynthia Z. F., op. cit.

TABLEAU 2. Taux de réponse pour la liste d'adresses du recensement de l'agriculture	1982	1978	1974
Adresses répertoriées	3,654,674	4,429,633	4,182,374
Questionnaires retournés par les services postaux (PMR)	82,792	230,980	108,700
Liste d'adresses – Exception faite des PMR	3,571,702	4,198,653	4,073,674
Non-réponse (y compris les nouveaux envois)	531,916	532,030	525,875
Questionnaires reçus	3,039,966	3,666,623	3,547,799
Exploitations visées	2,021,400	2,044,989	2,029,389
Exploitations non visées	978,264	1,511,218	1,487,351
Cas indéterminés	40,302	110,416	31,059
Ensemble des exploitations de la liste			
Taux de réponse global (%)	85.4	88.0	87.4
Exploitations classifiées (%)	82.1	80.3	84.1
Autres adresses (PMR et cas indéterminés) (%)	3.3	7.7	3.3
Non-réponses (%)	14.6	12.0	12.6
Exploitations de la liste exception faite des PMR			
Taux de réponse global (%)	85.1	87.3	87.1
Exploitations classifiées (%)	84.4	84.7	86.3
Autres adresses (%)	1.1	2.6	.8
Non-réponses (%)	14.9	12.7	12.9

Source: U.S. Bureau of the Census, données finales tirées des totalisations relatives au dénombrement postal fournies par l'Agriculture Division

vue d'avoir une idée plus précise de la fréquence optimale des suivis postaux[10].

Pour calculer les taux de réponse hebdomadaires pour la période allant du 21 janvier au 9 septembre 1983, on a divisé le nombre total de questionnaires renvoyés par le nombre total de questionnaires expédiés. Le nombre total de questionnaires renvoyés incluait tous les questionnaires renvoyés par la poste (remplis ou non), tout envoi provenant d'exploitants répertoriés sur la liste d'adresses ainsi que les questionnaires qui avaient été retournés par les services postaux parce qu'ils n'avaient pu être livrés. Les exploitations agricoles "probables" auxquelles un questionnaire avait été envoyé ont été réparties en 16 catégories selon le chiffre prévu de leurs ventes pour 1982. Ces catégories ont ensuite été ramenées à cinq groupes pour les besoins de l'étude. Les chiffres de ventes correspondant à chaque groupe sont respectivement: A – au moins $100,000, B – de $10,000 à $99,999, C – de $1,000 à $9,999, D – moins de $1,000 et E – indéterminé.

Environ 25% des exploitants inscrits sur la liste d'adresses ont reçu le questionnaire complet. Ce dernier reprenait toutes les questions du questionnaire abrégé auxquelles s'ajoutait un certain nombre de questions supplémentaires. Les exploitants auxquels le questionnaire complet était adressé appartenaient à deux groupes. Le groupe des "exploitations qui étaient incluses avec certitude dans l'échantillon" comprenait les exploitations qui, de par leur taille et le code

10 Ruggles, Donna R., et al., op. cit.

source, étaient considérées comme des exploitations agricoles de grande taille (chiffre des ventes prévu de l'ordre de $40,000 ou plus). Environ 328,000 exploitations répertoriées sur la liste d'adresses, soit 9%, se classaient dans ce groupe. Les autres exploitants qui ont reçu le questionnaire complet avaient été sélectionnés à partir des autres adresses de la liste. Ces derniers, au nombre de 573,000, constituaient l'échantillon "facultatif". Le reste des exploitants répertoriés sur la liste d'adresses (2.8 millions) avait reçu le questionnaire abrégé.

À la date limite du 15 février, le taux de réponse cumulatif à l'échelle nationale était de 46%. Ce taux a atteint un point culminant entre janvier et la mi-avril, a décru jusqu'à la mi-juillet et s'est stabilisé ensuite jusqu'à la fin de la collecte des données (figure 1). Chacun des six suivis postaux visant les non-répondants a permis d'améliorer le taux de réponse (tableau 1). Seuls le deuxième et le sixième de ces suivis postaux comprenaient un questionnaire et une lettre de rappel. L'augmentation la plus marquée du taux de réponse hebdomadaire est survenue après le deuxième suivi postal; il s'agit d'une augmentation de 12.8% enregistrée entre le deuxième et le troisième suivis (figure 2). Trois semaines après le sixième suivi, le taux de réponse était supérieur de 2.7% à ce qu'il avait été durant les dix semaines précédentes. Cela revient à dire que l'expédition du questionnaire contribue plus sûrement que le simple envoi d'une lettre à régler les cas de non-réponse.

Dans le cas des groupes d'exploitations agricoles "probables" définis selon la taille, exception faite de celui correspondant aux exploitations dont la taille n'avait pu être déterminée, les taux de réponse ont crû de façon similaire de janvier à la mi-juin (figure 3). Les suivis postaux ont eu sensiblement le même effet quel que soit le groupe. Le suivi téléphonique visant les non-répondants du groupe A, qui a débuté à la mi-mai, a résulté en une augmentation plus importante du taux de réponse cumulatif pour ce groupe que pour les autres groupes. À la fin juin, le groupe A avait le plus haut taux de réponse cumulatif de tous les groupes et ce taux a continué de croître jusqu'à la fin de la collecte des données, au point que le taux de réponse final a été de 97.7%. Le taux particulièrement élevé atteint dans le cas de ce groupe illustre l'efficacité du suivi téléphonique.

Les taux de réponse cumulatifs des groupes B, C et D ont augmenté de façon analogue de la mi-juin à la fin de la période de calcul des totalisations. Dans le cas du groupe E, le taux de réponse final a été nettement inférieur à celui des groupes correspondant aux exploitations dont la taille avait pu être déterminée. La proportion d'exploitations agricoles "véritables" parmi les répondants était de beaucoup inférieure dans le cas de ce groupe par rapport aux autres groupes. Le faible taux de réponse enregistré pour ce groupe particulier peut être dû à une imprécision des instructions relatives à qui devait remplir le questionnaire.

La longueur du questionnaire ne semble pas avoir eu d'incidence sur les taux de réponse (figure 4). Toutefois, en juin, quand le suivi téléphonique a commencé pour les exploitations incluses avec certitude dans l'échantillon auquel était destiné le questionnaire complet, le taux de réponse pour la portion du groupe visé par le questionnaire complet a augmenté. Les taux de réponse pour l'échantillon facultatif visé par le questionnaire complet et pour le groupe d'exploitations visé par le questionnaire abrégé étaient sensiblement analogues. Cela prouve que l'accroissement du fardeau de réponse occasionné par le questionnaire complet n'a pas eu un effet négatif sur le taux de réponse.

À partir de cette évaluation des taux de réponse, on peut tirer plusieurs conclusions. Les suivis postaux comportant un questionnaire en plus de la lettre de rappel se sont avérés plus efficaces pour améliorer le taux de réponse que les suivis postaux comportant uniquement la lettre de rappel. Cela a été vérifié à l'occasion de l'évaluation des variations du suivi postal. Cependant, le suivi téléphonique visant les non-répondants dont le chiffre des ventes était de $100,000 ou plus s'est avéré la méthode la plus efficace pour accroître le taux de réponse dans le cas de ce groupe particulier. La longueur du questionnaire complet ne semble pas avoir influé sur le taux de réponse. Le regroupement des destinataires du questionnaire de recensement selon le chiffre prévu de leurs ventes n'a révélé aucune corrélation précise entre la taille de l'exploitation et le taux de réponse.

4.3 Efficacité de la variation des procédures de suivi

L'objectif de l'évaluation des variations du suivi postal était de déterminer s'il y avait une

Figure 1. Taux de réponse cumulatif à l'échelle nationale

Réponse cumulative en pourcentage
100
90
80
70
60
50
40
30
20
10
0
28 janv. 25 févr. 18 mars 15 avr. 13 mai 10 juin 8 juill. 5 août 2 sept.
Date

Figure 2. Augmentation hebdomadaire du taux de réponse cumulatif à l'échelle nationale

Augmentation en pourcentage
8
7
6
5
4
3
2
1
0
28 janv. 25 févr. 18 mars 15 avr. 13 mai 10 juin 8 juill. 5 août 2 sept.
Date

Figure 3. Taux de réponse cumulatifs selon la taille des exploitations

Figure 4. Taux de réponse cumulatifs selon le type d'échantillon

différence statistiquement significative dans le taux de réponse postale selon que l'envoi comportait un questionnaire ou seulement une lettre de rappel[11]. Pour les besoins de l'évaluation, on a sélectionné un groupe d'essai et un groupe de contrôle. La procédure adoptée pour l'envoi initial et pour le premier suivi était identique à celle qui avait été utilisée pour le recensement. Pour les deuxième et troisième suivis, les envois destinés au groupe d'essai comprenaient d'abord une lettre de rappel et ensuite un questionnaire, soit l'inverse des envois correspondants effectués lors du recensement et des envois destinés au groupe de contrôle.

Pour des raisons financières, la sélection de l'échantillon a été limitée à 13 états et la taille de l'échantillon à 100,000 unités. Les états où l'échantillon a été tiré ont été choisis parce qu'ils appartenaient à deux types de régions bien particulières du point de vue de la taille des exploitations. Sept d'entre eux (Virginie, Caroline du Nord, Caroline du Sud, Géorgie, Kentucky, Tennessee et Alabama) étaient des états du Sud caractérisés par une majorité d'exploitations de petite taille (tant du point de vue de la superficie que de la valeur des produits) pour lesquels les taux de réponse aux recensements antérieurs avaient été les plus bas. Les six autres états (Ohio, Indiana, Illinois, Iowa, Nebraska et Kansas) avaient été sélectionnés dans la région du Midwest où l'on avait enregistré les taux de réponse les plus élevés aux recensements antérieurs. L'évaluation prenait en considération un autre élément, soit la longueur du questionnaire, c'est-à-dire questionnaire complet ou abrégé. La procédure d'échantillonnage systématique a abouti à la sélection d'environ 5,000 exploitants répertoriés sur la liste d'adresses pour chaque type de questionnaire et chaque groupe de chacun des 13 états visés.

Jusqu'à la première date repère (19 mars), le groupe d'essai et le groupe de contrôle avaient été traités de façon identique, c'est-à-dire qu'ils avaient reçu l'envoi initial avec le questionnaire et l'envoi avec la fiche de rappel. On n'avait alors relevé aucune différence dans les taux de réponse. Le 19 mars correspond au deuxième suivi à l'occasion duquel un questionnaire a été envoyé au groupe de contrôle alors que le groupe d'essai recevait la lettre de rappel. Une deuxième date repère (23 avril) a été choisie après l'envoi correspondant au troisième suivi. À cette date, les deux groupes avaient reçu un questionnaire et la lettre de rappel, mais dans un ordre inverse. Enfin, la dernière date repère (21 mai) correspondait au quatrième suivi.

Une analyse à plusieurs variables du modèle de covariance a été effectuée dans laquelle le taux de réponse à la première date repère constituait la variable de liaison et les taux de réponse cumulatifs aux deuxième et troisième dates repères les variables dépendantes. Le modèle comportait trois facteurs: le groupe (groupe de contrôle et groupe d'essai), le questionnaire (complet et abrégé), la région (Sud et Midwest) et un élément d'interaction (type de questionnaire par région). L'analyse a révélé que le groupe et la région avaient un effet déterminant mais qu'il n'y avait pas de différence sensible attribuable au type de questionnaire ou à l'élément d'interaction. Par ailleurs, on a enregistré une différence moyenne entre les régions pour la variable de liaison qui s'est révélée être un élément significatif du modèle. La figure 5 illustre les résultats de cette analyse.

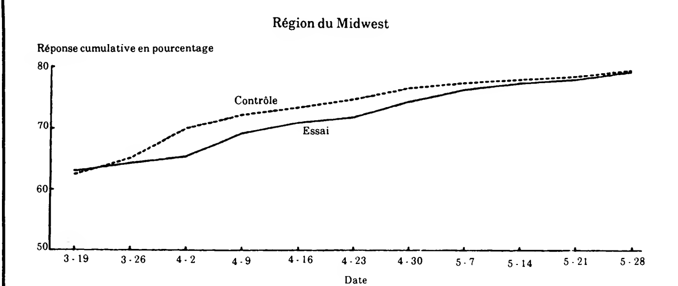
L'analyse du modèle de covariance a été reprise en utilisant les mêmes termes. Dans ce deuxième cas, la deuxième date repère (23 avril) a servi de variable de liaison et la troisième date repère (21 mai) de variable dépendante. On a obtenu les mêmes résultats qu'à la première analyse, soit une différence due au groupe et à la région mais aucune différence significative attribuable au questionnaire ou à l'élément d'interaction.

Les résultats de l'évaluation des variations du suivi postal indiquent clairement que les envois comportant un questionnaire sont plus efficaces que ceux qui comprennent seulement la lettre de rappel (tableau 3). Cette conclusion devra être prise en considération au moment de la planification du programme de suivi postal pour le recensement de l'agriculture de 1987. En vue d'améliorer ce programme, il conviendrait d'entreprendre une analyse des coûts portant sur le suivi postal, avec envoi d'un questionnaire, le plus coûteux et le suivi postal, avec simple envoi de la lettre de rappel, le moins coûteux. Un autre élément important de l'analyse de la rentabilité du suivi postal avec envoi d'un questionnaire est l'amélioration de la qualité des données due au fait que cette méthode permet d'obtenir plus rapidement un questionnaire rempli.

---

11 Ruggles, Donna R., et al., op. cit.

Figure 5. Taux de réponse cumulatifs par région utilisés pour l'évaluation des variations du suivi postal



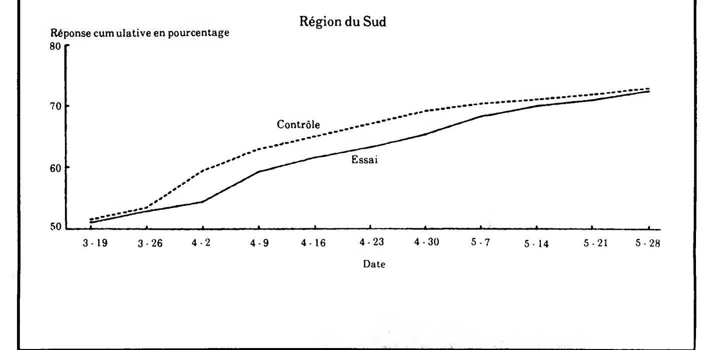


TABLEAU 3. Pourcentage de données du recensement de l'agriculture ayant été imputées	Liste d'adresses de 1982	Liste d'adresses seulement de 1978	Liste d'adresses et échantillon aréolaire de 1978	Liste d'adresses de 1974
NOMBRE D'EXPLOITATIONS PUBLIÉ				
Obtenu de la liste d'adresses (%)	90.2	82.5	90.7	87.7
Obtenu de l'échantillon aréolaire (%)	NA	8.9	NA	NA
Obtenu par imputation (%)	9.8	8.6	9.3	12.3
SUPERFICIE DES TERRES				
Données fournies par les recensés (%)	94.4	95.4	95.3	94.1
Données imputées (%)	4.6	4.6	4.7	5.9
CULTURES RÉCOLTÉES				
Données fournies par les recensés (%)	94.1	93.5	93.4	93.8
Données imputées (%)	5.9	6.5	6.6	6.2
VENTES DE PRODUITS AGRICOLES				
Données fournies par les recensés (%)	96.3	96.1	96.1	95.9
Données imputées (%)	3.7	3.9	3.9	4.1

Source: U.S. Bureau of the Census, Agriculture Division.

4.4 Évaluation du suivi téléphonique

Pour l'évaluation du suivi téléphonique visant les exploitants agricoles devant être dénombrés individuellement qui n'avaient pas retourné leur questionnaire rempli, on a tiré deux échantillons de 10,000 non-répondants une fois que le fichier des non-répondants a été créé. Ces échantillons ont été tirés à l'intérieur de chaque état par échantillonnage en grappes stratifié, la taille de l'échantillon pour chaque état étant établie en fonction du pourcentage de non-répondants admissibles pour cet état par rapport au chiffre total. À l'intérieur de chacune des strates constituées dans un état, on a alors tiré un échantillon systématique de paires de non-répondants, les éléments de chaque paire étant ensuite attribués au hasard à l'un des deux échantillons finals. Les exploitants de l'un de ces échantillons ont été interviewés selon la méthode de suivi téléphonique conventionnelle, par le personnel de la Data Preparation Division du Census Bureau, à partir de Jeffersonville, Indiana. Les exploitants constituant l'autre échantillon ont été dénombrés par interview téléphonique automatisée (CATI) à partir du Washington, D.C. Office du Census Bureau[12]. Pour ce faire, on a élaboré une version révisée du questionnaire du recensement de l'agriculture utilisé pour les interviews téléphoniques adaptée aux interviews téléphoniques automatisées (CATI).

L'analyse des résultats obtenus avec chaque méthode doit inclure une répartition des cas selon l'état final, les taux correspondant à chaque type d'état final, le nombre d'appels requis en moyenne par cas pour atteindre l'état final, le nombre total de minutes requis en moyenne par cas pour atteindre l'état final et le pourcentage d'éléments de données modifiés à l'occasion du contrôle complexe. Les résultats provisoires révèlent que les taux de réponse obtenus par interview téléphonique automatisée (CATI) sont légèrement supérieurs à ceux qui sont enregistrés avec la méthode de suivi téléphonique conventionnelle et que la qualité des données est relativement la même quelle que soit la méthode. Le rapport détaillé des résultats sera disponible à la fin de 1985[13].

4.5 Précision de la classification des réponses

Pour le calcul du taux de réponse final obtenu au recensement, les questionnaires retournés

[12] Ferrari, Pamela, Storm, Richard R. et Tolson, Francis D., "Computer-Assisted Telephone Interviewing in the 1982 Census of Agriculture",1984 **Proceedings of the Section on Survey Research Methods,** American Statistical Association.

[13] Ferrari, Pamela W., "An Evaluation of Computer-Assisted Telephone Interviewing used During the 1982 Census of Agriculture", U.S. Census Bureau Internal Report.

par les exploitants agricoles et par les recensés qui ne sont pas des exploitants agricoles, les cas non classifiés à l'issue du processus de collecte des données et les questionnaires non livrés retournés par les services postaux ont toujours été comptés dans les questionnaires reçus (tableau 2). Une mesure comparative de la classification des questionnaires retournés à la fin de la collecte des données pour les trois derniers recensements peut être dérivée de la répartition des questionnaires reçus[14]. Si l'on élimine les questionnaires non livrés retournés par les services postaux du nombre total des exploitants inscrits sur la liste d'adresses, le pourcentage de questionnaires qui avaient été classifiés à l'issue de la collecte des données était sensiblement le même aux recensements de 1982 et 1978 (84.4% et 84.7%). Cela tient à ce qu'au recensement de 1982 on a réussi à classifier durant la collecte des données une plus forte proportion de questionnaires reçus (exception faite des questionnaires non livrés retournés par les services postaux) qu'au recensement de 1978.

Une fois que l'on a reçu les questionnaires renvoyés par les exploitants inscrits sur la liste d'adresses, il faut les classifier. Les totalisations que nous avons décrites précédemment donnent une mesure des questionnaires reçus qui n'avaient pas été classifiés durant la collecte des données. L'évaluation de la couverture du recensement a fourni deux mesures de l'erreur de classification, l'une porte sur les exploitations agricoles véritables classées comme exploitations non agricoles (mal classées) et l'autre sur les exploitations non agricoles classées comme exploitations agricoles comptant divers types d'activités (surdénombrement). On peut comparer les pourcentages obtenus en divisant ces mesures de l'erreur de classification par le nombre estimatif d'exploitations agricoles visées (tableau 4). Pour ce qui est de la catégorie regroupant les exploitations agricoles dont le chiffre des ventes est de $2,500 ou plus, le nombre d'exploitations mal classées a baissé au cours de ces recensements. Par contre, le pourcentage estimatif d'exploitations comptées en trop a été plus élevé au recensement de 1982 qu'aux recensements antérieurs.

4.6 Imputation de données dans le cadre du recensement

Du fait que bon nombre d'exploitations figurant sur la liste d'adresses ne sont pas des exploitations agricoles, tous les non-répondants ne sont pas nécessairement des exploitants agricoles. Les données du recensement de l'agriculture sont donc ajustées pour faire état des exploitants agricoles non répondants; pour ce faire, on a estimé la proportion d'exploitants agricoles dans chaque état, on a imputé une valeur aux éléments de données pour ces non-répondants et on a inclus ces données imputées dans le calcul des estimations relatives à ces éléments de données destinées à être publiées. Comme les renseignements fournis par les recensés sont en général plus précis que les données imputées, le pourcentage de données obtenues directement des recensés constitue une mesure de la qualité des données[15]. Le pourcentage de cas où il a fallu imputer toutes les données du questionnaire s'est situé entre 9 et 10% (tableau 5) pour les deux derniers recensements, alors qu'il était de 12.3% en 1974. Comme nous l'avons souligné précédemment, si le taux de réponse était supérieur en 1974, par contre la proportion de questionnaires pour lesquels il a fallu imputer des données était aussi supérieure aux proportions correspondantes enregistrées aux recensements de 1978 et de 1982.

La proportion de données obtenues directement des recensés a toujours été supérieure dans le cas des autres éléments de données principaux – superficie de l'exploitation, cultures récoltées et valeur des ventes de produits agricoles – que dans celui du nombre d'exploitations publié. Du fait que, pour le recensement de l'agriculture, un suivi particulièrement intensif est prévu dans le cas des non-répondants qui exploitent des fermes pour lesquelles le chiffre de ventes prévu est élevé, la plupart des exploitations agricoles pour lesquelles les données doivent être imputées sont des exploitations de petite taille. En conséquence, ces données imputées ont une faible incidence sur les chiffres relatifs à la superficie des exploitations, aux cultures récoltées et à la valeur des ventes de produits agricoles. Le fait que la méthode utilisée pour imputer des données aux questionnaires des exploitants agricoles non répondants se prête bien au calcul des estimations est un facteur qui favorise la qualité des données publiées. Une étude portant sur d'autres méthodes possibles d'imputation des données sur l'ensemble des activités des exploitations agricoles doit être réalisée avant 1987.

[14] Clark, Cynthia Z.F., op. cit.

[15] Clark, Cynthia Z.F., op. cit.

TABLEAU 4. Estimations utilisées pour l'évaluation de la couverture du recensement de l'agriculture

ESTIMATION DE LA COUVERTURE

ÉTATS-UNIS
Toutes les exploitations agricoles

	Liste d'adresses de 1982	Liste d'adresses de 1978*	Liste d'adresses et échantillon aréolaire de 1978**	Liste d'adresses de 1974
Visées par le recensement	2,240,964	2,257,775	2,478,612	2,314,013
Non visées par le recensement	259,944	169,997	43,664	266,495
Mal classées	76,554	58,144	57,408	
Comptées en trop	113,623	15,456	23,962	41,026
Chiffre estimatif de l'univers	2,463,839	2,470,460	2,555,722	2,539,482
Couverture nette	222,875	212,685	77,110	225,469

ÉTATS-UNIS
Toutes les exploitations agricoles

	Liste d'adresses de 1982	Liste d'adresses de 1978*	Liste d'adresses et échantillon aréolaire de 1978**	Liste d'adresses de 1974
Visées par le recensement	91.0	91.4	97.0	91.1
Non visées par le recensement	10.6	6.9	1.7	10.5
Mal classées	3.1	2.4	2.2	
Comptées en trop	4.6	0.6	0.9	1.6
Chiffre estimatif de l'univers	100.0	100.0	100.0	100.0
Couverture nette	9.0	8.6	3.0	8.9

É.-U.
VTP $2.500

	Liste d'adresses de 1982	Liste d'adresses de 1978	Liste d'adresses et échantillon aréolaire de 1978	Liste d'adresses de 1974
Visées par le recensement	536,327	460,535	611,623	616,728
Non visées par le recensement	213,864	94,588	20,562	164,213
Mal classées	51,119	17,664	17,664	
Comptées en trop	50,624	2,208	3,048	8,482
Chiffre estimatif de l'univers	750,686	570,579	646,801	772,459
Couverture nette	214,359	110,044	35,178	155,731

É.-U.
VTP $2.500

	Liste d'adresses de 1982	Liste d'adresses de 1978	Liste d'adresses et échantillon aréolaire de 1978	Liste d'adresses de 1974
Visées par le recensement	71.4	.80.7	94.6	79.8
Non visées par le recensement	28.5	16.6	3.2	21.3
Mal classées	6.8	3.1	2.7	
Comptées en trop	6.7	0.4	0.5	1.1
Chiffre estimatif de l'univers	100.0	100.0	100.0	100.0
Couverture nette	28.6	19.3	5.4	20.2

É.-U.
VTP $2,500

	Liste d'adresses de 1982	Liste d'adresses de 1978	Liste d'adresses et échantillon aréolaire de 1978	Liste d'adresses de 1974
Visées par le recensement	1,702,973	1,794,938	1,864,687	1,695,047
Non visées par le recensement	46,080	75,409	23,102	102,282
Mal classées	25,435	40,480	39,744	
Comptées en trop	62,999	13,248	20,914	32,544
Chiffre estimatif de l'univers	1,711,489	1,897,579	1,906,619	1,764,785
Couverture nette	8,516	102,641	41,932	69,738
É.-U. Exploitations agricoles spéciales	1,664	2,302	2,302	2,238

É.-U.
VTP $2,500

	Liste d'adresses de 1982	Liste d'adresses de 1978	Liste d'adresses et échantillon aréolaire de 1978	Liste d'adresses de 1974
Visées par le recensement	99.5	94.6	97.8	96.0
Non visées par le recensement	2.7	4.0	1.2	5.8
Mal classées	1.5	2.1	2.1	
Comptées en trop	3.7	0.7	1.1	1.8
Chiffre estimatif de l'univers	100.0	100.0	100.0	100.0
Couverture nette	0.5	5.4	2.2	4.0

* Liste d'adresses de 1978: estimstions de 1978 dérivées du recensement.
** Liste d'adresses et échantillon aréolaire de 1978: estimations de 1978 dérivées de la liste d'adresses du recensement de l'échantillon aréolaire.

Source: U.S. Bureau of the Census, Agriculture Division, Coverage Evaluation Estimates.

TABLEAU 5. Évaluation du contact postal pour le recensement de l'agriculture de 1987

Date d'envoi	Date de réception	Suivi Plan A	Suivi Plan B	Suivi Plan C	Suivi Plan D	Taux de réponse prévu
12/01	12/15	-	-	Fiche pré-censitaire	Fiche pré-censitaire	0
12/15	01/01	Envoi postal initial	Envoi postal initial	Envoi postal initial	Envoi postal initial	0
01/10	01/22	Fiche de remerciement et de rappel	-	-	Fiche de remerciement et de rappel	0
02/05	02/10	Questionnaire	Questionnaire	Questionnaire	Questionnaire	50
03/07	03/12	Fiche et lettre	Fiche et lettre	Fiche et lettre	Fiche et lettre	68
03/29	04/03	Questionnaire	Questionnaire	Questionnaire	Questionnaire	75

4.7 Couverture de l'univers des exploitations agricoles

En vue de produire une mesure indépendante du nombre d'exploitations agricoles non comptées dans les données publiées du recensement, chaque recensement de l'agriculture fait l'objet d'une évaluation de la couverture depuis 1945[16]. Les échantillons utilisés à cette fin, pour 1978 et 1982, ont été conçus de façon à fournir, pour plusieurs composantes de la couverture du recensement, des estimations au niveau régional au lieu d'estimations au seul niveau d'état comme c'était le cas en 1974.

Des estimations du nombre total d'exploitations agricoles sont fournies dans la publication relative à l'évaluation de la couverture dans laquelle le "nombre estimatif d'exploitations agricoles" est le résultat de la somme des exploitations "visées par le recensement", des exploitations "non visées par le recensement" et des exploitations "mal classées" moins les exploitations "comptées en trop" Les estimations détaillées correspondantes ont été calculées[17] pour trois catégories, à savoir l'ensemble des exploitations agricoles, les exploitations dont le chiffre des ventes de produits agricoles est inférieur à $2,500 (exploitations de petite taille) et les exploitations dont le chiffre des ventes de produits agricoles est de $2,500 ou plus. Le tableau 4 donne les chiffres estimatifs relatifs à ces composantes pour les trois recensements qui nous intéressent. Afin de pouvoir établir des comparaisons avec les estimations de la couverture pour 1982 et 1974, des estimations distinctes ont été calculées en ce qui a trait aux exploitations inscrites sur la liste d'adresses de 1978. Les chiffres estimatifs publiés dans le cadre de l'évaluation de la couverture du recensement de 1978 ont été conçus en fonction de la double base de sondage adoptée pour ce recensement (c'est-à-dire la liste d'adresses et l'échantillon aréolaire).

Le tableau 4 présente chacune des composantes de la couverture en pourcentage des estimations de la couverture pour l'ensemble du recensement. Pour la période visée, l'estimation de l'échantillon relativement au pourcentage d'exploitations "visées par le recensement" pour l'ensemble des exploitations et pour les exploitations de petite taille est supérieure dans le cas du recensement de 1978 à double base. Cependant, pour ce qui est des exploitations agricoles dont le chiffre des ventes est de

16 Davie, William C., Lorenzen, Emily, Prochaska, D. Dean, "Coverage Evaluation for the 1982 Census of Agriculture", 1984 **Proceedings of the Section on Survey Research Methods**, American Statistical Association.

17 Clark, Cynthia Z.F., op. cit.

$2,500 ou plus, l'estimation de l'échantillon de 1982 relativement au pourcentage d'exploitations "visées par le recensement" était supérieure à celle du recensement de 1978 à double base. Par contre, l'estimation du pourcentage d'exploitations "non visées par le recensement" (c'est-à-dire ne faisant partie ni de la liste d'adresses ni de l'échantillon aréolaire de 1978) était de beaucoup inférieure, pour toutes les catégories d'exploitations, dans le cas du recensement de 1978 à double base par rapport aux autres recensements. Les estimations de l'erreur de classification, c'est-à-dire les exploitations agricoles classées comme exploitations non agricoles (mal classées) et exploitations non agricoles classées comme exploitations agricoles (comptées en trop), étaient plus élevées pour 1982 que pour 1978, pour toutes les catégories à l'exception des exploitations mal classées dont le chiffre des ventes était de $2,500 ou plus.

Pour comparer l'incidence relative de ces composantes de la couverture, on utilise une mesure appelée couverture nette qui est égale au nombre d'exploitations "non visées par le recensement" plus le nombre d'exploitations agricoles "mal classées" moins le nombre d'exploitations "comptées en trop". Dans le cas des estimations du recensement de 1978 à double base, la couverture nette exprimée en pourcentage était moindre pour l'ensemble des exploitations agricoles (3.0%) et de beaucoup inférieure dans le cas des exploitations de petite taille (5.4%). Toutefois, en raison d'une estimation relativement plus élevée du surdénombrement en 1982, dans le cas des exploitations agricoles dont le chiffre des ventes était de $2,500 ou plus, la couverture nette exprimée en pourcentage était moindre (0.5%) que pour tous les autres recensements.

5. Mesures à prendre en vue du prochain recensement

5.1 Essai précensitaire

Un essai précensitaire est une occasion unique d'évaluer de nouvelles méthodes ou procédures applicables à un programme de recensement. Ces dernières peuvent porter, entre autres choses, sur la présentation du questionnaire (c'est-à-dire la couleur, la présentation, la taille, etc.), sur le libellé des questions, sur l'utilisation de documents connexes, sur les procédures de suivi ou de contact postal, sur le fardeau de réponse et sur les attitudes influant sur le degré de participation des recensés. Pour le recensement de l'agriculture des États-Unis, un essai sera effectué au début de 1986. Parmi les objectifs visés, citons notamment:

1. évaluer l'incidence sur le taux de réponse de deux formats de questionnaire, un questionnaire-brochure et un questionnaire constitué d'une seule feuille avec rabats semblable à celui qui avait été utilisé en 1982;
2. estimer le fardeau de réponse occasionné par le questionnaire;
3. évaluer l'efficacité de différents types de contacts et suivis postaux;
4. évaluer d'autres libellés et présentations possibles des questions de 1982 ainsi que de toutes nouvelles questions proposées et mettre en évidence les raisons pour lesquelles tels éléments de questions ont été laissés en blanc;
5. mesurer la qualité et l'intégralité des réponses fournies;
6. évaluer la possibilité d'élargir la portée du tri des questionnaires préalable à l'introduction par clavier;
7. évaluer une nouvelle procédure d'introduction des données par clavier;
8. évaluer les modifications apportées aux programmes de contrôle informatisé.

Il est prévu d'utiliser une combinaison de plusieurs méthodes d'essai pour atteindre ces objectifs. Environ 45,000 questionnaires doivent être postés à la fin décembre 1985 et retournés au plus tard le 1er février 1986. Plusieurs nouvelles procédures de suivi postal seront mises à l'essai durant la collecte des données qui s'étendra sur les quatre premiers mois de 1985. Des interviews seront effectuées en mars dans plusieurs comtés sélectionnés dans des grappes distinctes. Celles-ci consisteront essentiellement à réinterviewer les recensés pour obtenir des renseignements relativement à la qualité et à l'intégralité des réponses recueillies. En mars également, des expériences en salles de cours et des études d'observation seront effectuées dans les régions géographiques générales correspondant aux comtés sélectionnés dans les grappes. Les essais relatifs au dépouillement des données seront faits au bureau de Jeffersonville, Indiana, où seront

acheminés les questionnaires remplis. Les données d'un échantillon de questionnaires seront introduites par clavier de façon à mesurer le taux de réponse par question et à évaluer les modifications apportées au contrôle informatisé du questionnaire.

L'échantillon postal pour le recensement a été conçu de façon à permettre une évaluation statistique de l'incidence sur le taux de réponse du type de questionnaire, brochure ou formule avec rabats, le taux de réponse global servant de mesure de cette incidence. Le fardeau de réponse apparent, pour chaque type de questionnaire, sera évalué au moyen de diverses séances d'observation de groupes en salle de cours. Les estimations du fardeau de réponse seront établies à partir de plusieurs sources. Le temps nécessaire pour remplir le questionnaire sera mesuré à l'occasion des expériences faites en salle de cours et des études d'observation. Toutefois, cette façon de faire ne tient pas compte du temps requis pour rassembler les renseignements demandés dans le questionnaire. Une estimation du temps nécessaire pour remplir le questionnaire sera également obtenue au moment de la réinterview en fonction des indications fournies par les répondants. On pourra alors, à partir de ces estimations du temps requis, comparer les fardeaux de réponse correspondant à chaque type de questionnaire.

À l'occasion de la même évaluation, on analysera également l'efficacité de plusieurs plans de contacts postaux. Ces plans varieront du point de vue du nombre et du contenu des envois postaux, mais non de l'intervalle entre les envois. Les envois peuvent contenir une lettre de présentation accompagnée d'une fiche faisant part des statistiques tirées du recensement, une fiche de remerciement et de rappel, des lettres de suivi accompagnées ou non du questionnaire et des fiches de suivi. L'échantillon a également été prévu de façon à pouvoir effectuer une évaluation distincte d'un groupe d'enregistrements qui ne correspondent probablement pas à des exploitations visées par le recensement. La taille de l'échantillon est suffisamment importante pour déceler un écart de 2% dans les taux de réponse selon les plans évalués, le seuil de signification étant de .05.

Les expériences effectuées en salle de cours et les études d'observation serviront également à évaluer de nouveaux libellés et de nouvelles façons de présenter les questions de 1982 en plus des questions entièrement nouvelles proposées pour le prochain recensement. Les expériences en salle de cours permettront de donner une répartition aléatoire des résultats relatifs aux diverses alternatives envisagées. Les études d'observation portant sur une solution à la fois permettront de mettre en évidence les problèmes que pourraient occasionner certains éléments de données au niveau de la réponse à fournir ou de la collecte des données. La qualité et l'intégralité des données recueillies dans les questionnaires seront évaluées à partir des taux de réponse relatifs à chaque élément de donnée obtenus à l'occasion de l'évaluation postale et des expériences faites en salle de cours. Pour certains éléments de données, la qualité et l'intégralité des données seront également mesurées à l'occasion des réinterviews.

Deux nouvelles méthodes de tri des questionnaires basées sur l'introduction par clavier d'un nombre différent de questionnaires seront évaluées au moyen de plusieurs essais mettant en évidence le temps requis, effectués au centre de dépouillement des données de Jeffersonville, Indiana. On comparera le nombre de frappes effectuées et le taux d'erreur enregistré, selon chacune des deux méthodes, à l'occasion de l'introduction par clavier des données de questionnaires provenant d'onze régions géographiques différentes. Les données introduites permettront d'évaluer les modifications envisagées relativement aux programmes de contrôle informatisé.

5.2 Recherche concernant le taux de réponse

Toute amélioration de la qualité des données du recensement de l'agriculture repose essentiellement sur un taux de réponse élevé. L'obtention d'un taux de réponse satisfaisant lorsque la collecte des données est faite par la poste pose un problème qu'il faut régulièrement réévaluer. Un projet de recherche portant sur les divers facteurs qui influent sur le taux de réponse obtenu au recensement de l'agriculture a été entrepris il y a deux ans. Jusqu'ici, il a permis de mettre en évidence certaines procédures utilisées qui ont une incidence sur le taux de réponse et a conduit à proposer que des études soient effectuées en vue d'établir de quelle manière ces procédures doivent être modifiées en vue d'améliorer le taux de réponse. L'objectif de ce projet de recherche est d'améliorer le taux de réponse pour tous les recensés, c'est-à-dire les exploitants agricoles aussi bien que les autres

recensés. Cela permettrait de réduire la proportion des exploitations de la liste d'adresses pour lesquelles des données doivent être imputées et, par le fait même, d'améliorer la qualité des données du recensement.

Le contenu et la planification des envois postaux peuvent aussi influer sur le taux de réponse. À l'occasion de la préparation du recensement d'essai, tous les documents destinés à être envoyés par la poste ont été passés en revue. Les envois postaux ont été adressés principalement à des exploitants agricoles. Cependant, bon nombre de recensés ne sont pas ou ne se considèrent pas comme des exploitants agricoles. Il importe par conséquent d'arriver à obtenir un questionnaire rempli de tout recensé qu'il soit ou non un exploitant agricole. C'est l'objectif que l'on s'est donné au moment de l'élaboration des lettres, fiches et instructions devant servir pour le recensement d'essai.

La publicité relative au recensement peut viser à régler certains problèmes liés au taux de réponse. Nous sommes en train d'analyser l'évolution des taux de réponse dans différentes régions géographiques durant toute la période de collecte des données. Des critères sont élaborés pour les comtés où le taux de réponse est faible. Nous nous proposons d'étudier les caractéristiques des exploitants agricoles ainsi que les caractéristiques démographiques des résidents de ces comtés. Nous élaborerons ensuite une publicité appropriée en vue d'améliorer le taux de réponse dans ces régions.

Nous sommes également en train d'évaluer les taux de réponse correspondant aux exploitants dont les noms ont été obtenus de sources particulières et aux exploitants dont le chiffre des ventes prévu se situe dans certaines limites. L'utilité des adresses fournies et le taux de réponse diffèrent considérablement selon les diverses sources à partir desquelles la liste d'adresses est élaborée. Des critères particuliers seront établis pour les sources dont on a tiré les exploitations agricoles et pour les exploitations agricoles de certaines tailles qui se caractérisent par un faible taux de réponse. Ensuite, des procédures de suivi postal et téléphonique seront élaborées visant les non-répondants de ces catégories en complément des plans généraux de suivi postal mis à l'essai.

Une étude des raisons spécifiques de la non-réponse au recensement de l'agriculture est actuellement en cours d'élaboration. Il est préférable d'effectuer ce type d'étude pendant le recensement proprement dit de façon à obtenir des réponses des répondants et des non-répondants sans décalage temporel. Dans cette perspective, la tenue d'une enquête téléphonique en même temps que le recensement de l'agriculture de 1987 est actuellement prévue. Cette enquête a été conçue de façon à fournir des projections estimatives du taux de réponse, en plus des raisons spécifiques à l'origine de la non-réponse, et à mettre en évidence les circonstances où il conviendrait d'appliquer des procédures différentes.

PROPOSITION RELATIVE AU CHOIX DE L'EMPLACEMENT DES TERRES COMME BASE DU RECENSEMENT DE L'AGRICULTURE

G. OLIVER CODE

DIVISION DE L'AGRICULTURE ET DES RESSOURCES NATURELLES
STATISTIQUE CANADA

Introduction

Le présent document expose le concept d'un recensement de l'agriculture basé sur l'emplacement des terres plutôt que sur celui du siège de l'exploitation, comme c'était le cas au Canada jusqu'à présent. La proposition vise principalement les parties du recensement relatives aux grandes cultures et à l'utilisation des terres et fait état des besoins particuliers du personnel de la Section des cultures, de Statistique Canada. Compte tenu des avantages que présente l'utilisation de l'emplacement des terres comme base de recensement, il est recommandé que l'on adopte la double approche suggérée pour le recensement de l'agriculture de 1991. Il est également recommandé d'adopter une nouvelle procédure pour la collecte des données relatives aux grandes cultures, les avantages et les inconvénients de cette nouvelle approche étant exposés par la même occasion. La dernière section du présent document décrit la recherche opérationnelle qu'il est nécessaire d'effectuer en vue de la mise en place de ce type de changement pour le recensement de 1991.

Choix du siège de l'exploitation comme base de recensement

Le choix du siège de l'exploitation agricole comme base de recensement signifie que toutes les données fournies sur le questionnaire de recensement sont attribuées à l'emplacement géographique du terrain qui a été désigné par l'exploitant agricole comme étant le siège de l'exploitation. Autrement dit, toute partie de l'exploitation située en dehors des limites géographiques du siège est traitée comme si elle était située à l'intérieur de ces limites.

Limites du choix du siège de l'exploitation comme base du recensement

Le siège de l'exploitation a été choisi comme base de la collecte des données parce que c'est une façon simple et commode de procéder qui simplifie la collecte et la compilation des données. Cependant, l'application de cette méthode peut entraîner des erreurs au niveau de la répartition géographique de certaines terres. Antérieurement, la distorsion qui en résultait était généralement minime du fait que les exploitations étaient de taille relativement petite et que leurs terres étaient plutôt regroupées. Le risque de distorsion s'est accru ces dernières années sous l'effet de l'adoption de modes de propriété ou de location des terres plus avantageux et plus variés et de l'accroissement de la taille des exploitations dont les terres ne sont pas nécessairement contiguës.

Le personnel chargé du recensement de l'agriculture a procédé à une étude de cas portant sur l'incidence de la distorsion géographique, due au choix de l'emplacement du siège de l'exploitation comme base de recensement, sur les données du recensement de 1981 se rapportant à la subdivision de recensement unifiée 410, en Saskatchewan. Il a été conclu que la distorsion mesurée dans le cas d'une subdivision de recensement unifiée typique était significative[1]. La distorsion était d'au moins +10% des terres déclarées, au niveau de la SRU, et atteignait -44% au niveau du secteur de dénombrement. Un secteur de dénombrement est la région que doit dénombrer un recenseur pendant le recensement. Les limites d'un secteur de dénombrement ne doivent pas traverser les limites municipales ni celles des circonscriptions électorales fédérales. Une subdivision de recensement unifiée est composée d'un ensemble de secteurs de dénombrement et constitue habituellement la plus petite unité le plus couramment utilisée dans les totalisations du recensement de l'agriculture.

En vue d'évaluer l'ampleur du problème, le personnel chargé du recensement de l'agriculture a comparé dernièrement la superficie totale des exploitations agricoles tirée du recensement de 1981 à la superficie totale réelle des terres au niveau des subdivisions de recensement unifiées. L'étude portait essentiellement sur la région des Prairies qui regroupe plus de 80% de la superficie totale des terres améliorées au Canada. Elle a révélé que la superficie globale des terres

[1] Burroughs, R. J., "Investigation into the Impact of Geographic Distortion Due to the Headquarters Rule", Ottawa, Statistique Canada, février 1983.

agricoles attribuées sur la base de l'emplacement du siège de l'exploitation aux subdivisions de recensement unifiées de Saskatchewan dépassait de 19% la superficie totale possible. C'est pour la Saskatchewan que l'écart est le plus grand, soit 24%. Ces chiffres donnent une idée de la gravité de la distorsion. Il faudrait effectuer une recherche plus approfondie pour mieux cerner l'étendue du problème.

Depuis 1966, la taille moyenne des exploitations agricoles en Saskatchewan s'est accrue de près de 15 acres par année. Cette tendance à un accroissement de la taille des exploitations combinée à une plus grande dispersion des terres constituant l'exploitation est continue et, logiquement, le risque de distorsion tend également à croître. L'existence de contrats de gérance, de location ou d'association complexes alliée au fait que nombre d'exploitants sont non résidents[2] est à l'origine de situations difficiles et souvent confuses du point de vue de la couverture se traduisant par des cas de chevauchement ou de sous-déclaration, par exemple. Il peut aussi arriver qu'il soit particulièrement difficile de désigner un emplacement comme siège de l'exploitation. Tant que l'on conservera l'emplacement du siège de l'exploitation comme base pour la collecte des données du recensement de l'agriculture, de tels problèmes seront inévitables.

Le personnel chargé du recensement de l'agriculture a reconnu l'existence du problème de distorsion relatif à l'emplacement des terres et a prévu un certain nombre de procédures visant à résoudre les cas les plus flagrants et les plus graves. Ainsi, le personnel communique directement avec les responsables des exploitations de grande taille constituées en corporation afin de déterminer l'emplacement exact des terres de chaque exploitation; il vérifie les enregistrements relatifs aux exploitants non résidents et examine le cas de toute terre agricole située dans une région urbaine importante. Toutes ces précautions n'empêchent pas qu'il y ait des incohérences sérieuses au niveau des données de base relatives aux terres dans certaines subdivisions de recensement unifiées.

Les utilisateurs veulent avoir des données régionales plus fiables et plus flexibles. Il y a une demande croissante pour une plus grande variété de données permettant d'analyser les questions en rapport avec l'économie, l'environnement, les ressources, la gestion et la production agricoles. Les statisticiens agricoles provinciaux ont souligné que les données régionales tirées du recensement sont inadéquates pour bon nombre des tâches de cet ordre et, ce qui est plus grave, contiennent des erreurs qui peuvent conduire à tirer des conclusions fausses des analyses.

La Section des cultures de la Division de l'agriculture et des ressources naturelles produit une série statistique sur la superficie, le rendement et la production des principales céréales. Dans le cadre de la compilation de ces statistiques, on utilise les données régionales pour la pondération et la sélection de l'échantillon en ce qui a trait aux grandes cultures. Il est essentiel d'avoir une base aréolaire précise pour pouvoir produire des estimations aléatoires valables.

Méthode recommandée - Choix de l'emplacement des terres comme base de recensement

En ce qui a trait aux données sur les grandes cultures et sur l'utilisation des terres, il conviendrait d'adopter l'emplacement des terres comme base du recensement de l'agriculture. Cela garantirait une plus grande homogénéité des données relatives à l'identification spatiale et à l'utilisation des terres. L'unité de déclaration serait un emplacement déterminé et la collecte des données porterait sur les caractéristiques de terres géographiquement définies au lieu de se rapporter à l'exploitation.

Avantages du choix de l'emplacement des terres comme base de recensement

Par opposition aux inconvénients de plus en plus marqués liés à l'utilisation de l'emplacement du siège de l'exploitation comme base de la collecte des données, le choix de l'emplacement des terres offre un certain nombre d'avantages du fait que cela permet de rendre compte de l'emplacement exact des grandes cultures.

Ce type de données serait très utile du point de vue du programme de Statistique Canada relatif aux grandes cultures. Avec une base de sondage complète il serait possible de revenir au concept de la base aréolaire, c'est-à-dire de tirer un échantillon de terres au lieu de tirer un échantillon d'exploitants à partir d'une liste. Cette façon de faire serait en outre parfaitement adaptée au programme de télédétection pour lequel on ne peut employer que des données basées sur l'emplacement des terres. Avec ce type de données il devient plus faisable, d'un point de vue opérationnel, d'utiliser les données recueillies

[2] Un exploitant non résident est un exploitant agricole qui demeure en dehors des terres de son exploitation.

grâce au programme de télédétection pour produire des statistiques sur les superficies et, éventuellement, sur le rendement, ce qui permettrait d'alléger le fardeau de réponse des exploitants agricoles. Compte tenu des progrès prévus en ce qui a trait à l'utilisation des satellites, il devrait être possible de suivre l'évolution annuelle des superficies et d'utiliser les données ainsi obtenues pour produire des données régionales intercensitaires de bonne qualité. La télédétection par satellite pourrait devenir d'ici 1996 une composante essentielle des opérations de collecte de données sur les grandes cultures.

La couverture globale du recensement serait améliorée si l'on pouvait utiliser des procédures de recensement basées sur l'emplacement des terres qui soient plus faciles à définir. Par exemple, les problèmes liés aux déclarations incomplètes et aux cas de chevauchement seraient alors plus apparents et plus faciles à résoudre. La taille de l'exploitation, le mode de gérance ou de location même complexe n'importerait plus du fait qu'il n'y aurait plus de référence à l'exploitant agricole.

Les utilisations possibles d'une banque de données basées sur l'emplacement des terres seraient variées, englobant des domaines aussi divers que l'environnement et l'utilisation des terres. Les possibilités de recouvrement des coûts seraient considérablement accrues car nous serions en mesure de répondre à des demandes nouvelles et différentes. La Direction générale des terres d'Environnement Canada a un programme de contrôle de l'utilisation des terres dont les données sont mises à jour tous les cinq ans. Le recensement permettrait d'alimenter ce projet.

Il existe plusieurs fichiers administratifs basés sur l'emplacement des terres (c'est-à-dire les cartes selon le type de sol, les données utilisées pour assurer les récoltes, les registres de permis de la Commission canadienne du blé). Ces fichiers pourraient être appariés aux données du recensement basées sur l'emplacement des terres afin de produire des ensembles de données entièrement nouveaux ou des séries portant sur des problèmes particuliers.

Différentes possibilités de mise en application de l'approche basée sur l'emplacement des terres

Diverses méthodes peuvent être utilisées en vue d'intégrer au recensement de l'agriculture l'approche basée sur l'emplacement des terres. En voici quelques-unes.

a) Aux États-Unis, la question suivante est posée au recensement de l'agriculture: Dans quel comté vous attendez-vous à avoir la plus grande valeur de cultures ou de produits? Autrement dit, l'attribution des terres dans le cadre du recensement américain est basée sur les ventes et non sur l'emplacement du siège de l'exploitation; toutefois, cette façon de faire occasionne des problèmes du même type que ceux que pose le système canadien.

b) Une application plus rigoureuse des méthodes actuellement utilisées par Statistique Canada pour attribuer les données sur l'agriculture permettrait de réduire le problème. Toutefois, les procédures correspondantes, que nous avons décrites dans la section relative aux limites du système actuel, n'apportent pas de solution véritable ni partielle aux problèmes propres à l'analyse des données régionales.

c) Les données relatives à l'emplacement géographique de chaque terre comprise dans une exploitation agricole pourraient être introduites en machine. Cela permettrait le repérage automatisé des secteurs problèmes et une certaine ré-attribution des terres. Là encore, il s'agit essentiellement d'un moyen de réduire la portée du problème et non d'une solution à long terme.

d) On pourrait élaborer une matrice de questionnaire permettant d'attribuer un identificateur unique à chaque type d'utilisation des terres et à l'emplacement correspondant déclarés dans le questionnaire de recensement. Cette méthode faciliterait la compilation d'une banque importante de données géocodées dont le format pourrait être rapidement modifié en fonction de presque n'importe quel critère de sélection. L'inconvénient de cette solution est que les questionnaires de ce type sont réputés difficiles à remplir avec exactitude.

Technique de collecte des données recommandée

Un des points forts du recensement de l'agriculture est la cohérence des questions et des techniques de collecte utilisées d'un recensement à l'autre. Pour préserver cet avantage, il est recommandé d'utiliser un double système de collecte des données en 1991.

Le système actuel, basé sur l'emplacement du siège de l'exploitation, devra être conservé pour toutes les variables traditionnelles, y compris les grandes cultures. Le recenseur complétera ces

données en demandant à l'exploitant de délimiter l'emplacement des diverses grandes cultures et d'en identifier le type. Il est trop tôt pour établir quel serait le système de déclaration des données le plus efficace et le moins onéreux mais ce serait probablement l'utilisation d'une photographie aérienne, d'une carte topographique ou d'un cliché obtenu par satellite. Les données pourraient ensuite être converties en données numériques et utilisées pour produire une base de données géographiques.

Aspects positifs de l'approche orientée vers la production de cartes

Ces aspects sont les suivants:

a) la précision du système actuel se trouvera accrue par suite de la nécessité de faire état sur une carte ou sur une photographie de toutes les terres situées dans une région donnée;

b) les exploitants agricoles semblent très favorables à l'emploi de photographies aériennes, ils prennent même un certain plaisir à ce mode de déclaration de sorte que le fait de devoir fournir ce supplément de données sera à peine perçu ou ne sera pas perçu du tout comme alourdissant le fardeau de réponse;

c) le personnel de la Division de l'agriculture et des ressources naturelles de Statistique Canada a beaucoup d'expérience dans l'utilisation des photographies aériennes qui sont utilisées avec succès pour l'enquête nationale sur les fermes;

d) le processus complexe de conversion des limites des grandes cultures en données numériques peut être réalisé au moyen des techniques existantes; toutefois, des techniques améliorées devraient être au point d'ici 1991;

e) dans la mesure où le module relatif aux grandes cultures est tiré du contexte usuel du recensement, l'actualité et la continuité des données de recensement publiées ne s'en trouveront pas altérées;

f) cette approche fait appel aux compétences du personnel en poste à Statistique Canada, notamment à l'expérience acquise par celui-ci en matière de manipulation de bases de données complexes. Le personnel de la Division de la géographie a une bonne expérience du traitement des données spatiales et pourrait avoir un rôle déterminant au niveau de la recherche de nouveaux systèmes;

g) pour le recensement de l'agriculture on utilise déjà des plans de township avec une grille spéciale pour les quarts de section pour assurer une meilleure couverture de la région des Prairies. Le recenseur doit inscrire le numéro du questionnaire dans chaque quart de section. Ce serait donc peu de choses de modifier le système et de demander au recenseur d'indiquer, à la place, les grandes cultures et l'utilisation des terres.

Limites du système proposé

L'adoption d'un système basé sur l'emplacement des terres et d'une méthode de collecte des données différente présente quelques inconvénients possibles.

a) Cette double approche entraînera un surcroît de dépenses. On ne peut encore déterminer jusqu'à quel point le coût lié à l'établissement par le recenseur d'un double module de questions sera de beaucoup supérieur au coût de l'établissement du questionnaire actuel. Il est difficile de déterminer avec précision, à ce stade-ci, le coût net d'une telle opération ou d'estimer les profits que l'on pourrait tirer du recouvrement des coûts.

b) Il y aura quelques différences entre les données sur les grandes cultures recueillies dans le cadre du recensement traditionnel et celles qui auront été obtenues grâce au système cartographique. Ces différences devront être expliquées au public, encore que la nouvelle méthode confirmera très probablement l'exactitude globale du système actuel.

c) Avec le nouveau système, le dépouillement, l'agrégation et la production des données prendront plus de temps. La rapidité d'exécution dépendra des ressources allouées.

d) Du fait que cette nouvelle technique est complètement différente de ce qui est en usage actuellement, il y aura nécessairement des problèmes opérationnels. Il faudra procéder à des essais sur le terrain et à des essais opérationnels d'envergure.

e) La mise sur pied d'un système d'analyse et d'une banque de données fondées sur l'emplacement des terres nécessite un ordinateur puissant.

f) Dans la mesure où il y aurait une demande pour la répartition selon l'emplacement des terres des données se rapportant à d'autres variables du recensement, il faudrait faire des recherches et concevoir un modèle approprié.

Recherche concernant les problèmes opérationnels

La recherche nécessaire pour pouvoir mettre sur pied un tel système peut être présentée sous quatre rubriques:

(1) Collecte des données: quel est le meilleur système géographique à utiliser; quels systèmes de codage et d'identification doivent être employés; quelles sont les améliorations les plus avancées en matière d'analyse de données et de matériel de conversion numérique; essai sur le terrain.

(2) Couplage des données: quel est le système de saisie des données le plus efficace; comment sera effectuée la conversion numérique des données; quelle procédure de dépouillement des données doit être employée; quel système de vérification des données doit être adopté.

(3) Stockage des données: quel système de stockage des données doit être adopté; quel type de manipulation électronique doit être utilisé; quel est le volume de stockage des données informatiques à prévoir.

(4) Diffusion des données: quel niveau de désagrégation doit être prévu; quelles sont les variables de données les plus importantes; quel type de systèmes d'accès est à prévoir pour les utilisateurs; quels sont les problèmes qui se posent du point de vue de la confidentialité des données.

CONFIDENTIALITÉ DES DONNÉES DU RECENSEMENT DE L'AGRICULTURE

RICK BURROUGHS

DIVISION DE L'AGRICULTURE ET DES RESSOURCES NATURELLES
STATISTIQUE CANADA

MARY MARCH

DIVISION DES MÉTHODES DE RECENSEMENT ET D'ENQUÊTES-MÉNAGES
STATISTIQUE CANADA

Introduction

Un certain nombre d'articles de la Loi sur la statistique traite de la question de la confidentialité des données de recensement. L'alinéa 16(1)b) en particulier s'applique au recensement de l'agriculture et stipule qu': "aucune personne qui a été assermentée en vertu de l'article 6 ne doit révéler ni sciemment faire révéler, par quelque moyen, des renseignements obtenus en vertu de la présente loi de manière qu'il soit possible, grâce à de telles révélations, de rattacher à un particulier, à une entreprise ou à une organisation identifiables les détails obtenus dans un relevé qui les concerne exclusivement". L'application de ce principe à une statistique donnée, considérée individuellement, suppose que l'analyste a une connaissance précise des données à partir desquelles cette dernière est dérivée et dispose de tout le temps nécessaire pour en évaluer l'importance. Toutefois, il est plus difficile d'appliquer ce même principe dans le cas d'une base de données importante comportant de multiples variables et donnant lieu à une foule de possibilités d'extraction.

Pour ce qui est d'une telle base de données, la meilleure façon de procéder est sans doute la suivante:

a) assurer que les exigences relatives à la confidentialité des données, telles qu'elles sont énoncées dans la Loi sur la statistique, sont respectées dans **tous** les cas de diffusion des données. En d'autres mots, aucun communiqué de presse, publication, totalisation spéciale ni fichier de données remis à des utilisateurs de l'extérieur de Statistique Canada ne doit fournir les caractéristiques précises d'une exploitation agricole ou d'un exploitant agricole (ou permettre de découvrir de telles caractéristiques par recoupement avec des données diffusées antérieurement);

b) optimiser le volume et la qualité des données rendues disponibles. En d'autres mots, les ambiguïtés insérées en vue de préserver la confidentialité des données doivent être évitées de façon à sauvegarder l'utilité des estimations fournies;

c) permettre la mise sur pied de la base de données à un coût raisonnable.

Nature de la base de données

Si l'on veut évaluer avec justesse le problème de confidentialité propre au recensement de l'agriculture, il est essentiel d'avoir une bonne connaissance de la nature de la base de données. Nous nous servirons du recensement de l'agriculture de 1981 comme exemple car, du fait qu'il s'agit du recensement le plus récent, c'est certainement la base de données qui est le plus susceptible de s'apparenter à celle du recensement de 1991. Le nombre d'exploitations étant à la baisse et cette tendance risquant de se maintenir, les problèmes qui se sont posés au recensement de 1981 seront probablement plus marqués en 1991.

La base de données du recensement de 1981 comprend quelque 318,000 enregistrements (exploitations agricoles visées par le recensement). Chaque enregistrement peut compter jusqu'à 278 zones de données ou variables. Cependant, toutes les variables ne s'appliquent pas nécessairement à chaque exploitation agricole visée; ainsi, une variable comme le nombre de poulets sera assortie d'une valeur uniquement dans les enregistrements (exploitations agricoles visées par le recensement) qui font état de poulets. Par contre, une variable telle que l'âge de l'exploitant devra être assortie d'une valeur dans tous les enregistrements.

Par conséquent, les enregistrements contiennent en général moins de 278 zones. Un enregistrement moyen contient 14 zones avec

indicateurs numériques et géographiques, 18 zones avec des données qualitatives[1] et 42 zones avec des données quantitatives[2], soit au total 74 zones.

Les résultats normalisés fournis à partir d'un enregistrement incluent une répartition géographique de toutes les variables pour la province, la division de recensement et la subdivision de recensement unifiée, et sept classements recoupés des principales variables au niveau de la province. Toutes les autres statistiques sont extraites à la demande des utilisateurs. Le Bureau a reçu plus de 1,000 demandes de ce genre depuis le début de la diffusion de la base des données de 1981 en juillet 1982.

Règle des dix exploitations

La procédure le plus couramment utilisée aux derniers recensements est la Règle des dix exploitations. Selon cette règle, aucune donnée ne peut être diffusée concernant une région géographique qui compte moins de 10 exploitations agricoles, les données doivent être combinées à celles de régions adjacentes jusqu'à ce que l'ensemble des régions ainsi regroupées donne un total de 10 exploitations ou plus. Le tableau 1 illustre l'application de cette règle au moyen d'un ensemble de données fictives.

L'importance de cette règle peut être illustrée en mesurant son effet sur les niveaux géographiques normalisés qui sont utilisés dans le cadre du recensement. Sur les 257 divisions de recensement au Canada qui comptent des exploitations agricoles, quatre contiennent moins de 10 exploitations; sur les 2,475 subdivisions de recensement unifiées (SRU) qui comptent des exploitations agricoles, 307 contiennent moins de 10 exploitations et plus de la moitié des 13,000 secteurs de dénombrement qui comptent des exploitations agricoles contiennent moins de 10 exploitations.

Cette règle est relativement peu coûteuse à appliquer et généralement, quoique pas toujours, bien acceptée par les utilisateurs. Bien qu'elle permette d'éviter bon nombre de révélations illicites, elle n'empêche pas certaines révélations malencontreuses et, par le fait même, interdit inutilement la diffusion de statistiques acceptables. Ainsi, dans le tableau 1, le nombre de porcs pour la SRU 3 est sujet à révélation illicite et exempt de diffusion, dans le cas de la SRU 2 il est également sujet à révélation illicite mais non exempt de diffusion alors que, dans le cas de la SRU 1, il n'est nullement sujet à révélation illicite mais est exempt de diffusion inutilement.

Suppression de chiffres

La suppression de chiffres est une technique qui n'est plus utilisée actuellement mais qui l'était encore en 1976. Essentiellement, cela consiste à décider de ne pas diffuser le nombre d'exploitations pour lesquelles des données ont été enregistrées concernant une variable particulière, au-dessous d'un niveau géographique spécifique. Cette technique était généralement appliquée dans le cas de variables quantitatives relativement peu fréquentes, comme certaines cultures horticoles. Le tableau 2 illustre l'application de cette technique au moyen des mêmes données fictives que celles qui ont servi pour le tableau 1. Il a été décidé de ne pas diffuser le nombre d'exploitations ayant déclaré des porcs au niveau des SRU en raison de la forte incidence des révélations illicites à ce niveau. Dans l'exemple qui est donné, les données pour les SRU 2 et 3 pourraient donner lieu à des révélations illicites alors que ce n'est pas le cas pour les données relatives à la SRU 1.

Cette technique est, elle aussi, d'application peu coûteuse et permet d'éliminer les données susceptibles de faire l'objet de révélations illicites. Dans la plupart des cas, les chiffres supprimés ne sont pas eux-mêmes objets de révélations illicites, de sorte que le coût en termes de perte de données est relativement élevé. L'inconvénient majeur de cette technique est qu'elle n'élimine pas les révélations illicites, elle les rend seulement plus difficiles à découvrir.

Règle de 3 appliquée aux distributions

La règle de 3 appliquée aux distributions est une autre technique qui n'a plus cours mais qui était encore utilisée au recensement de 1976. Elle est appliquée aux distributions de fréquence des variables quantitatives. De façon générale, elle s'énonce comme suit: lorsque le total pour la catégorie la plus élevée contenant une valeur est 3

[1] Le terme "qualitative" décrit ici une variable composée de codes classant l'exploitation selon un attribut ou une qualité en particulier, par exemple, le type d'exploitation.

[2] Le terme "quantitative" décrit ici une variable constituée d'une somme ou d'un volume d'éléments en rapport avec l'exploitation, par exemple le nombre de porcs.

TABLEAU 1. Règle des "dix exploitations"

SANS LA RÈGLE DES DIX EXPLOITATIONS

	Nombre d'exploitations	Total des ventes (en milliers de $)	Nombre de porcs	Nombre d'exploitants ayant déclaré des porcs
SRU 1	6	27	34	4
SRU 2	15	105	1,250	2
SRU 3	7	22	4	1
Division de recensement	28	154	1,288	7

AVEC LA RÈGLE DES DIX EXPLOITATIONS

	Nombre d'exploitations	Total des ventes (en milliers de $)	Nombre de porcs	Nombre d'exploitants ayant déclaré des porcs
SRU 1[1]	–	–	–	–
SRU 2	15	105	1,250	2
SRU 3[1]	13	49	38	5
Division de recensement	28	154	1,288	7

1 Six exploitations de la SRU 1 ont été combinées avec celles de la SRU 3.

TABLEAU 2. Règle de la "suppression de chiffres"

SANS LA SUPPRESSION DE CHIFFRES

	Nombre d'exploitations	Total des ventes (en milliers de $)	Nombre de porcs	Nombre d'exploitants ayant déclaré des porcs
SRU 1	6	27	34	4
SRU 2	15	105	1,250	2
SRU 3	7	22	4	1
Division de recensement	28	154	1,288	7

AVEC LA SUPPRESSION DES CHIFFRES

	Nombre d'exploitations	Total des ventes (en milliers de $)	Nombre de porcs	Nombre d'exploitants ayant déclaré des porcs
SRU 1	6	27	34	–
SRU 2	15	105	1,250	–
SRU 3	7	22	4	–
Division de recensement	28	154	1,288	7

et que celui pour la catégorie suivante la plus basse est 0, on combine alors les deux catégories les plus élevées contenant une valeur puis on remplace l'échelle de valeurs pour cette catégorie par "égal ou supérieur à la valeur minimale résultant de la combinaison des deux catégories". Cette technique est aussi appliquée au total d'une variable associée. Le tableau 3 illustre cette procédure à l'aide d'une distribution de fréquences du chiffre total des ventes tiré des données relatives à la division de recensement utilisée comme exemple dans les tableaux 1 et 2.

Cette fois encore, il s'agit d'une technique d'application peu coûteuse. Celle-ci est probablement très efficace dans le cas des catégories supérieures des distributions de fréquences, particulièrement lorsqu'on lui adjoint un ensemble soigneusement sélectionné de limites de catégories. Elle ne permet pas de régler les problèmes se rapportant aux catégories intermédiaires ou inférieures pour lesquelles les révélations illicites sont réputées non identifiables ou insignifiantes.

Suppression de totaux

Cette technique visant à éviter les révélations illicites a été utilisée, sous diverses formes, lors de plusieurs recensements de l'agriculture, bien que le recensement de 1981 soit le premier où elle ait été utilisée de façon systématique.

La procédure consiste essentiellement à supprimer les totaux relatifs à des variables quantitatives à chaque fois que le nombre d'exploitations en cause est petit. La suppression d'un total entraîne habituellement celle d'un ou plusieurs autres totaux complémentaires de façon à prévenir toute révélation illicite de données connexes. Les règles d'aplatissement servant à définir un tel total varient de façon significative. Les deux règles adoptées pour le recensement de 1981 sont connues sous les noms de "règle Duffett" et de "règle de moins de 3 unités déclarantes".

Le tableau 4 reprend la division de recensement fictive utilisée dans les tableaux précédents pour illustrer l'incidence de cette technique sur les données.

Après un examen comparatif des tableaux 3 et 4, on peut être porté à penser qu'il y a dans la SRU 2 une seule exploitation agricole pratiquant l'élevage de porcs, dont le chiffre des ventes est de $87,000. Supposons que cela est exact. Il s'ensuit que les totaux relatifs au nombre de porcs et au chiffre de ventes, au niveau de la SRU 2, et au nombre de porcs, au niveau de la division de recensement, sont sujets à révélations illicites. Ils devront donc être combinés à d'autres totaux correspondant à d'autres divisions de recensement. Le nombre de porcs dans la SRU 3 est, bien entendu, sujet à révélations illicites et le nombre de porcs dans la SRU 1 de même que le chiffre des ventes dans la SRU 3 sont des compléments.

Ce type de procédure s'avère généralement efficace. Le principal inconvénient qui lui est lié est l'importance des ressources humaines aussi bien qu'informatiques que sa mise en application requiert. L'élaboration du logiciel CONFID, par le Bureau, est une approche systématique et détaillée qui a permis d'analyser, pour la première fois en 1981, les problèmes qui peuvent se poser du point de vue de la confidentialité des données.

Cependant, les coûts impliqués ont limité l'analyse à 64 variables quantitatives sur 246 et aux seuls produits normalisés. Pour ce qui est des données produites à la demande des utilisateurs, on a appliqué, manuellement, la règle de moins de 3 unités déclarantes à ces mêmes 64 variables.

Cette méthode présente un autre inconvénient en termes de perte de données. Le nombre de compléments insérés dans la majorité des données extraites dépasse le nombre de données sujettes à des révélations illicites et croît en proportion géométrique avec le nombre d'entrées du tableau. Cela est particulièrement malencontreux lorsqu'un total insignifiant mais sujet à des révélations illicites entraîne une chaîne de compléments dont la somme dépasse largement le total en question.

Arrondissement aléatoire à la base 5

L'arrondissement aléatoire à la base 5 est une technique qui a été appliquée aux bases de données de 1971 et de 1981 pour le couplage agriculture-population.

Il s'agissait d'arrondir tout nombre d'exploitations agricoles extrait de la base au multiple de 5 le plus proche, de façon aléatoire ou au moyen d'un coefficient de probabilité déterminé. Toutes les autres données associées au nombre ainsi révisé sont ensuite ajustées de la même manière de façon à préserver l'exactitude des moyennes. Le tableau 5 montre l'incidence de cette règle sur les données relatives à la division de recensement utilisée comme exemple dans les tableaux précédents.

L'arrondissement aléatoire n'entraîne aucun coût dans la mesure où l'on utilise pour ce faire la version actuellement disponible du logiciel

TABLEAU 3. Règle de "3 appliquée aux distributions"

SANS LA RÈGLE DE 3 APPLIQUÉE AUX DISTRIBUTIONS

DIVISION DE RECENSEMENT

Exploitations dont le total des ventes est de l'ordre de:	Nombre d'exploitations	Total des ventes (en milliers de $)
Moins de $1,000	11	8
$1,000 à $4,999	12	35
$5,000 à $9,999	4	24
$10,000 à $24,999	0	0
$25,000 et plus	1	87
Total	28	154

AVEC LA RÈGLE DE 3 APPLIQUÉE AUX DISTRIBUTIONS

Exploitations dont le total des ventes est de l'ordre de:	Nombre d'exploitations	Total des ventes (en milliers de $)
Moins de $1,000	11	8
$1,000 à $4,999	12	35
$5,000 et plus	5	111
Total	28	154

TABLEAU 4. Règle de la "suppression de totaux"

SANS LA SUPPRESSION DE TOTAUX

	Nombre d'exploitations	Total des ventes (en milliers de $)	Nombre de porcs	Nombre d'exploitants ayant déclaré des porcs
SRU 1	6	27	34	4
SRU 2	15	105	1,250	2
SRU 3	7	22	4	1
Division de recensement	28	154	1,288	7

AVEC LA SUPPRESSION DES TOTAUX

	Nombre d'exploitations	Total des ventes (en milliers de $)	Nombre de porcs	Nombre d'exploitants ayant déclaré des porcs
SRU 1	6	27	–	4
SRU 2	15	–	–	2
SRU 3	7	–	–	1
Division de recensement	28	154	–	7

TABLEAU 5. Règle de la "l'arrondissement aléatoire"

SANS L'ARRONDISSEMENT ALÉATOIRE

	Nombre d'exploitations	Total des ventes (en milliers de $)	Nombre de porcs	Nombre d'exploitants ayant déclaré des porcs
SRU 1	6	27	34	4
SRU 2	15	105	1,250	2
SRU 3	7	22	4	1
Division de recensement	28	154	1,288	7

AVEC L'ARRONDISSEMENT ALÉATOIRE

	Nombre d'exploitations	Total des ventes (en milliers de $)	Nombre de porcs	Nombre d'exploitants ayant déclaré des porcs
SRU 1	10	45	43	5
SRU 2	15	105	0	0
SRU 3	5	16	20	5
Division de recensement	25	138	1,840	10

STATPAK-TARELA. Il présente en outre l'avantage que les données ainsi produites ne risquent pas de donner lieu à des révélations illicites. Par contre, cette procédure comporte un inconvénient regrettable du fait que, malgré les apparences, elle ne règle nullement les problèmes de dominance.

Ces particularités apparaissent clairement dans le tableau 5. L'exploitation d'élevage porcin n'est plus aussi évidente une fois que les données ont été arrondies mais son existence peut néanmoins être déduite à partir du total relatif à la division de recensement. Ce tableau illustre un autre problème, celui de la distorsion des données brutes. Une telle distorsion est significative uniquement lorsque le nombre d'exploitations est peu élevé, ce qui est souvent le cas avec les données sur l'agriculture, et de nombreuses utilisations des données deviendront impossibles si cette procédure est généralisée.

Procédures relatives à la confidentialité pour le recensement de 1986

En 1986, le principal objectif sera de prévenir le mieux possible les révélations illicites compte tenu des contraintes financières. Étant donné que les ressources disponibles en ce qui a trait aux systèmes sont peu importantes, les procédures de prévention des révélations illicites devront être limitées à l'utilisation de CONFID ou de l'arrondissement aléatoire actuellement disponibles. L'élaboration de nouveaux systèmes devra être limitée aux systèmes nécessaires pour assurer une interface valable entre le logiciel relatif à la confidentialité des données et les systèmes propres au recensement de l'agriculture.

Les procédures relatives à la confidentialité s'appliqueront à tous les produits du recensement de 1986, y compris les totalisations et les fichiers de microdonnées compilés à la demande des utilisateurs. Ces procédures incluront des mesures destinées à éviter les révélations illicites touchant des données connexes par suite de la combinaison des tableaux produits sur demande avec d'autres données disponibles. Le Bureau cherchera à réduire les risques de révélations illicites par un usage judicieux des procédures disponibles.

Méthodes proposées pour le recensement de 1991

Aucune des procédures relatives à la confidentialité utilisées jusqu'à présent, pas même l'approche envisagée pour 1986, ne sont véritablement satisfaisantes. Le logiciel CONFID lui-même, qui permet d'éviter la publication de données sujettes à des révélations illicites, élimine parfois des données qui semblent nécessaires (il est difficile d'accepter que des estimations non sujettes à des révélations illicites

portant sur des produits importants soient éliminées pour éviter de révéler des renseignements touchant un exploitant particulier qui est le seul à déclarer ce produit, en petites quantités, dans sa région). En 1981, la mise sur pied de CONFID s'est avérée une tâche considérable qu'il a fallu insérer dans un programme de production très serré. Aucune procédure n'est aussi efficace que l'arrondissement aléatoire lorsqu'il s'agit d'éviter les révélations illicites de données connexes qui pourraient être occasionnées par la publication de données sur demande.

Cette procédure de prévention des révélations illicites est une composante importante du processus d'enquête. Par conséquent, tous les efforts nécessaires doivent être entrepris pour assurer qu'elle est aussi efficace que les autres composantes de l'opération et donne des résultats de qualité comparable.

Dans le cadre de la planification du recensement de 1991, il conviendrait d'envisager la possibilité d'adopter d'autres méthodes. Il pourrait s'agir tout simplement de reprendre des procédures qui ont déjà été utilisées, en les modifiant. Par exemple, l'arrondissement aléatoire peut parfaitement s'appliquer à certaines variables se caractérisant par des distributions sans surprise et pourrait même être très efficace moyennant un certain contrôle permettant d'éviter les distorsions excessives tout en introduisant suffisamment d'ambiguïté. La procédure CONFID est, elle aussi, une solution attrayante car elle évite les révélations illicites, particulièrement dans le cas de plusieurs produits ayant des dénominateurs communs. Il suffirait de l'employer de manière sélective. Une toute nouvelle approche pourrait également être envisagée. Par exemple, on pourrait étudier la possibilité d'adopter une procédure s'appliquant, au niveau des microdonnées, à tous les enregistrements ou à certains seulement. Le problème posé par la confidentialité des données se situe, en fait, au niveau des microdonnées puisqu'il s'agit de protéger les données se rapportant à des **particuliers**. Pourquoi n'adopterait-on pas une méthode basée sur la distorsion ou la combinaison des microdonnées?

Quelle que soit l'approche adoptée pour 1991, il faudrait qu'elle puisse être reprise aux recensements subséquents. En d'autres mots, il conviendrait de mettre un terme à la remise en cause permanente et aux modifications incessantes en ce qui concerne les procédures relatives à la confidentialité des données recueillies dans le cadre du recensement de l'agriculture.

Procédures de prévention des révélations illicites proposées pour les futurs recensements de l'agriculture

La confidentialité des données est une des préoccupations majeures de Statistique Canada. Dans la pratique, un groupe de travail spécial a été mis sur pied pour étudier les procédures actuellement appliquées et faire des recommandations pour l'avenir (concernant la normalisation, les mécanismes à mettre en place, etc.). À la suite des travaux de ce groupe et d'autres types d'interventions, des lignes directrices seront vraisemblablement élaborées qui pourront être appliquées directement au recensement de l'agriculture.

Si des mécanismes appropriés et commodes devenaient applicables dans l'intervalle, en ce qui a trait à la prévention des révélations illicites, il sera possible de les utiliser, à condition, bien sûr, que le mécanisme en question soit parfaitement adapté au recensement de l'agriculture dont les données diffèrent, sous plusieurs rapports, des autres produits de Statistique Canada.

En réalité, même si l'on adoptait une solution "toute faite", celle-ci devrait subir certaines modifications. L'utilisation des programmes de combinaison et de suppression de données du CONFID pour le recensement de 1981 a révélé que l'insertion d'un mécanisme tout prêt relatif à la confidentialité des données peut s'avérer une tâche considérable, particulièrement lorsque le nombre d'estimations en cause est important et que les interfaces avec les autres parties du système causent des difficultés. Une analyse doit être effectuée pour assurer que les procédures relatives à la confidentialité peuvent être intégrées au système de production de données. Idéalement, ces procédures devraient être aussi efficacement intégrées au système que c'est le cas pour l'imputation, par exemple.

Une étude minutieuse du contexte du recensement de l'agriculture est nécessaire pour déterminer le type de procédure le plus susceptible d'assurer la confidentialité des données. Le type de révélations illicites possibles dépend de la répartition géographique et de la répartition par taille des données ainsi que du type de tableau produit (y compris le nombre d'entrées, le nombre de niveaux pour chaque entrée, la présence de chiffres de fréquences pour les variables quantitatives, le fait que les catégories sont ou non cumulatives, etc.). Le choix des procédures utilisées pour prévenir les révélations illicites doit également être fondé sur les utilisations qui seront faites des données.

Il ne faut pas perdre de vue que toutes les diffusions de données, pour un recensement particulier, ne sont pas simultanées et ne sont pas nécessairement prédéterminées. Les fichiers et les totalisations produits sur demande doivent être traités adéquatement pour éviter toute révélation illicite de données connexes par suite de leur utilisation avec d'autres données publiées.

Il est recommandé de débuter les travaux d'analyse et de développement bien avant 1991 si l'on veut pouvoir mettre en oeuvre une procédure efficace à cette époque.

Description et calendrier des activités de développement proposées pour 1991

Si l'on veut pouvoir incorporer au recensement de 1991 une procédure relative à la confidentialité des données qui soit efficace tant du point de vue du coût que de la qualité, deux activités doivent avoir été menées à terme avant que l'équipe chargée du recensement de 1991 entreprenne l'élaboration de spécifications pour les systèmes du recensement (c'est-à-dire environ trois ans avant le jour du recensement).

La première activité, soit l'analyse des situations pouvant donner lieu à des révélations illicites et des exigences qualitatives exprimées par les utilisateurs, devra être terminée au plus tôt. Une analyse partielle a déjà été faite dans le cadre du développement des systèmes pour les recensements de 1981 et 1986. Des analyses plus approfondies devront être effectuées par les spécialistes du domaine essentiellement (avec la collaboration de méthodologistes pour l'étude des distributions de données, par exemple). Cette tâche devra absolument être terminée avant juin 1987 (ce qui laisse une année pour les activités consécutives à cette recherche avant le début du processus d'élaboration des spécifications pour 1991).

La seconde activité, soit l'incorporation d'une procédure appropriée aux méthodes adoptées pour le recensement de l'agriculture, devra être terminée l'année suivante. Idéalement, cette procédure devrait être intégrée au système du recensement au même titre que l'imputation. De fait, elle pourrait même faire partie intégrante du processus d'imputation de sorte que nous aurions trois bases: la base de données non imputées, la base de données imputées et la base de données ajustées pour préserver la confidentialité[3]. Il est sans doute plus réaliste d'envisager que la procédure relative à la confidentialité sera incorporée au programme de production.

---

[3] Cela suppose qu'une procédure visant les microdonnées est appropriée. Il y a imputation de données pour les cas de non-réponse, il y aurait également imputation de données dans le cadre de la procédure relative à la confidentialité (c'est-à-dire remplacement ou déplacement des données fournies).

Calendrier de développement proposé

Activité	Date
1. Définition des besoins du recensement de l'agriculture en matière de système relatif à la confidentialité.	Juin 1987
2. Mise au point de la méthode.	Juin 1988

Résumé de la discussion

M. Jones, chargé du recensement de l'agriculture de 1986, a repris le contenu des trois exposés.

Méthode d'envoi et retour du questionnaire par la poste

Le premier exposé, qui portait sur le recensement de l'agriculture aux États-Unis, présente un intérêt particulier pour cet auditoire du fait qu'il donne une idée des problèmes qui risquent de se poser à l'occasion du recensement de l'agriculture au Canada si l'on adopte la méthode d'envoi et retour du questionnaire par la poste. Il conviendrait de ne pas perdre de vue les avantages liés à la combinaison du recensement de l'agriculture avec celui de la population. En plus des économies que cela permet de réaliser, il faut rappeler que le recensement de l'agriculture étant de taille minime par rapport au recensement de la population, son budget est moins facilement visé par des coupures sévères. Il serait peu sage de mettre fin à la relation qui lie les deux recensements.

Les similarités entre le recensement des États-Unis et celui du Canada se manifestent souvent, en particulier, aux niveaux des problèmes liés aux concepts et définitions et des programmes publicitaires respectifs.

Proposition relative à l'adoption de l'emplacement des terres comme base du recensement de l'agriculture

Le deuxième exposé concernait l'adoption proposée de l'emplacement des terres comme base de la collecte des données du recensement de l'agriculture, avec utilisation des données obtenues par satellite. Le point faible d'une telle procédure est le risque que les appareils tombent en panne.

L'adoption de l'emplacement comme base du recensement a déjà été proposée par des utilisateurs d'Environnement Canada, d'Agriculture Canada, du ministère de l'Agriculture de Saskatchewan et le milieu de la recherche. Du fait qu'un des objectifs du recensement est de fournir des données régionales et du fait que la méthode actuelle de collecte des données entraine des distorsions aux niveaux géographiques inférieurs, il est évident qu'il faut effectuer des recherches en vue de mettre au point des méthodes mieux appropriées.

Les méthodes suggérées en vue d'avoir des données basées sur l'emplacement des terres entraînent une augmentation substantielle des coûts. En outre, en cherchant à augmenter la qualité des données régionales, comme il est proposé dans l'exposé, on risque d'inciter les utilisateurs à demander davantage de données détaillées pour un plus grand nombre de variables, ce qui ne ferait qu'accroître le coût du recensement. Même si une partie des frais peut être remboursée par les utilisateurs, le Bureau pourrait manquer de fonds pour faire face à la dépense supplémentaire.

Comme cela a été souligné dans l'exposé précédent relatif à l'élargissement de la méthode de retour du questionnaire par la poste, il faut réduire plutôt qu'augmenter les coûts sur le terrain des enquêtes. Toute mesure conduisant à accroître la complexité des procédures sur le terrain représente un pas en arrière dans le contexte actuel.

Procédures relatives à la confidentialité des données

Le troisième exposé portait sur les procédures relatives à la confidentialité des données du recensement de l'agriculture. Ces procédures ont été modifiées à chacun des trois derniers recensements et, malgré cela, on n'a toujours pas clairement établi ce que l'on entend par révélations illicites. Il est nécessaire de réduire aussi bien les risques de révélations illicites (quelle qu'en soit la définition) que le coût des procédures adoptées à cet effet.

À l'occasion de l'élaboration d'un nouveau système, il faut veiller à maintenir un certain équilibre entre la préservation de la confidentialité des données fournies par les répondants et la qualité des données mises à la disposition des utilisateurs.

Il serait souhaitable de pouvoir utiliser un système généralisé applicable à plusieurs enquêtes, mais cette perspective est encore lointaine. Une tentative d'élaboration d'un système spécialisé pour le recensement de l'agriculture peut aboutir à la prise en charge de cette tâche par la direction, ce qui libérerait des fonds pour le système généralisé. Quoi qu'il en soit, il serait utile d'entamer le développement du système relatif au recensement de l'agriculture.

Résumé

*En conclusion, **M. Jones** a souligné que le travail de recherche, pour ce qui est des deux domaines mentionnés, devrait commencer au plus tôt. En outre, il conviendrait d'effectuer des recherches en rapport avec l'élargissement proposé de la méthode de retour du questionnaire par la poste. En vue de faciliter le travail, il serait possible d'ajouter, dans les formules 2A et 2B du recensement de la population, une question visant les exploitants agricoles.*

Résumé de la période de questions

La séance relative au recensement de l'agriculture a été suivie par une période de questions au cours de laquelle les participants ont fait part de leurs préoccupations relativement aux points abordés pendant la séance.

Besoin de données régionales

Le besoin de données régionales a été mis en doute. De telles données sont-elles utiles en elles-mêmes ou parce que les utilisateurs veulent les utiliser pour faire leurs propres agrégations? S'agit-il d'un petit nombre d'agrégations? Auquel cas, ne serait-il pas plus simple de fournir ces agrégations plutôt que de produire des données régionales?

L'équipe chargée du recensement de l'agriculture aux États-Unis ne reçoit que très peu de demandes concernant les données régionales. Les plus petites régions visées par ces demandes sont généralement constituées de deux ou trois comtés. Les totalisations spéciales sont, en général, assez coûteuses (coût moyen de l'ordre de $50,000), de sorte qu'il n'y a eu que 100 demandes de ce genre en ce qui concerne les données du recensement de 1982.

Au Canada, il y a beaucoup de demandes de totalisations spéciales des données du recensement de l'agriculture. Ces totalisations sont relativement peu onéreuses (coût moyen de l'ordre de $300). Certains utilisateurs préfèrent effectuer les agrégations eux-mêmes, alors que d'autres demandent qu'on leur fournisse les totaux relatifs à des zones de préservation, à des secteurs de drainage, à un district fédéral, etc. Toutes les demandes sont traitées dans la mesure où elles n'enfreignent pas les exigences relatives à la confidentialité des données. Il y a une demande pour une plus grande flexibilité à ce chapitre.

Lien entre le recensement de l'agriculture et le recensement de la population

La question du lien entre le recensement de l'agriculture et celui de la population a été abordée. De toute évidence, le travail de planification devra être effectué dans la perspective du maintien de la combinaison des deux recensements. Toutefois, un changement de la relation qui unit les deux recensements peut s'avérer inévitable s'il devenait nécessaire d'éliminer le plus possible le recours à des intervieweurs pour le recensement de la population. On pourrait alors envisager la possibilité de limiter à la couverture le rôle du recensement de la population par rapport à celui de l'agriculture, ce dernier devenant en quelque sorte une enquête de suivi.

Confidentialité des données

Le reste de la discussion a porté sur la confidentialité des données. Le sentiment général est qu'il y a trop de données supprimées et que des procédures plus efficaces doivent être mises au point pour le traitement des demandes spéciales. Bien que les méthodologistes des divers pays commencent à échanger leurs connaissances concernant les procédures relatives à la confidentialité des données, ils n'en sont pas encore à travailler de concert. Il a été souligné que le United States Bureau of the Census et Statistique Canada sont toujours au stade de la recherche d'une meilleure coordination interne à cet égard. Il a été suggéré que le sens de "révélations illicites" adopté pour le recensement de l'agriculture soit compatible avec les définitions utilisées pour la diffusion de données commerciales.



SÉANCE: AUTOMATISATION DU RECENSEMENT

Président: Martin Podehl
Division des services et du
développement informatiques
Statistique Canada

Jeudi, le 10 octobre 1985

PLANS D'AUTOMATISATION DU RECENSEMENT DE LA POPULATION ET DU LOGEMENT DE 1990 DES ÉTATS-UNIS

PETER A. BOUNPANE

U.S. BUREAU OF THE CENSUS

Introduction

L'automatisation est l'une des principales questions à l'étude dans le cadre de la planification du recensement de la population et du logement de 1990 des États-Unis. Automatiser bon nombre des tâches manuelles du recensement de 1980 et des recensements précédents nous permettra d'accélérer le dénombrement tout en remplissant notre mandat juridique qui consiste à diffuser les chiffres de la représentation électorale et de la répartition des districts, et d'autres résultats dans les meilleurs délais. L'automatisation nous aiderait aussi à rentabiliser de nombreux secteurs, à améliorer la précision de nos résultats de même qu'à mieux contrôler le processus de recensement.

Depuis toujours, la collecte des données du recensement et la plupart des traitements qu'on leur fait subir (par ex., le contrôle à l'arrivée des questionnaires par rapport à une liste de contrôle d'adresses, le contrôle d'intégralité des questionnaires et le codage des réponses écrites) sont des tâches qui exigent beaucoup de paperasses et de travail. L'utilisation de matériel automatisé permet d'aborder le flot de paperasserie et la multitude de travaux d'écritures d'une façon beaucoup plus efficace et contrôlée. Jusqu'ici, recruter et former les employés nécessaires à l'exécution des nombreuses tâches du recensement, et leur trouver des locaux prenaient beaucoup de temps et coûtaient cher. Même si le recensement de 1990 nécessitera vraisemblablement l'embauchage d'un grand nombre de travailleurs temporaires, nous cherchons des moyens de réduire le nombre d'activités à forte intensité de main-d'oeuvre et d'utiliser des systèmes automatisés pour contrôler le processus de recensement.

Nous travaillons sur des plans d'automatisation depuis quelque temps déjà. Nous avons testé de nouvelles méthodes lors des recensements d'essai que nous avons effectués cette année à Tampa en Floride et à Jersey City dans le New Jersey et nous mènerons d'autres essais d'automatisation l'année prochaine dans certaines parties du comté de Los Angeles en Californie et dans plusieurs comtés de East Central Mississippi. Ces essais sont très importants puisqu'ils constituent des laboratoires où nous pouvons mettre à l'épreuve diverses méthodes et j'y reviendrai tout au long de ce document.

Étant donné que nous envisageons des changements considérables dans l'automatisation en vue du recensement de 1990, je vais d'abord décrire comment le recensement de 1980 s'est déroulé pour que vous puissiez mieux comprendre en quoi consistent ces changements.

Recensement de 1980

En 1980, les questionnaires du recensement ont été envoyés et retournés par la poste dans les régions du pays qui regroupaient 95% de la population. Nous avons acheté les listes d'adresses de certaines régions et dressé ces listes nous-mêmes ailleurs; dans les deux cas, les listes ont été vérifiées et mises à jour par le U.S. Postal Service et par notre propre personnel sur le terrain. Le USPS a livré des questionnaires à tous les logements quelques jours avant le recensement et les membres responsables des ménages devaient les remplir et les renvoyer par la poste à un bureau de district de recensement temporaire le 1er avril. L'objet de cette méthode était de faire le plus gros du recensement par l'entremise du service postal qui revenait moins cher et de réserver les suivis coûteux et longs aux logements dont les occupants n'avaient pas retourné de questionnaire. Moins de deux semaines après l'envoi postal, nous avions reçu les questionnaires remplis d'environ 83% des ménages. Un important effectif (jusqu'à 270,000 employés à un certain moment) a visité les ménages non répondants et les logements inoccupés. Dans les régions du pays à faible densité de population où les procédures d'envoi et de retour par la poste des questionnaires ne convenaient pas, les recenseurs ont visité tous les ménages.

Nous avons mis sur pied provisoirement 409 bureaux de district pour réaliser la collecte des données. Dans tous les bureaux, un grand nombre de commis ont été embauchés pour apporter des changements (additions, suppressions et corrections) aux listes d'adresses, contrôler à l'arrivée les questionnaires renvoyés par la poste et leur faire subir un contrôle d'intégralité et d'uniformité, attribuer aux recenseurs les logements à suivre, surveiller le dénombrement des ménages non répondants et totaliser des chiffres provisoires. La plupart de ces opérations ont été faites manuellement. Bon nombre d'entre elles peuvent être considérées comme des "traitements", mais n'ont pas vraiment débuté avant

que les bureaux de collecte n'aient eu terminé leur travail, fermé leurs portes et expédié leurs questionnaires à l'un des trois centres de traitement. Dans l'ensemble, les bureaux ont fermé entre 5 et 7 mois après le jour de recensement.

Aux centres de traitement, les questionnaires ont été microfilmés et les données introduites en mémoire au moyen de lecteurs FOSDIC. Même si les opérations des centres de traitement étaient en grande partie automatisées, les réponses en toutes lettres à beaucoup de questions, notamment celles sur l'origine et sur la profession, ont dû être codées numériquement à la main avant d'être soumises aux traitements sur ordinateur. Ces opérations de codage ont nécessité un grand nombre de commis.

Ce système a très bien fonctionné, compte tenu du volume de travail manuel à faire et de la nette démarcation entre la collecte et le traitement des données. Le Census Bureau a respecté les délais prévus par la loi pour la diffusion des chiffres de la représentation électorale et de la répartition des districts; bon nombre des données régionales, par exemple, les chiffres d'îlot, ont été diffusées plus tôt qu'au recensement précédent et beaucoup d'autres données, en particulier sur les races et sur les groupes d'origine espagnole, ont été publiées. Cependant, nous n'avons pas diffusé aussi rapidement que prévu certains résultats, particulièrement ceux qui reposaient sur des questions-échantillon. (Ce retard était en partie attribuable à des problèmes budgétaires qui nous ont forcés à réduire le personnel et à suspendre momentanément les opérations de codage des échantillons.)

Pour le recensement de 1990, nous voulons de nouveau respecter les délais prescrits par la loi et diffuser d'autres résultats plus rapidement que jamais tout en faisant en sorte que les coûts restent raisonnables et que les chiffres soient aussi précis que possible.

Plans d'automatisation pour 1990

Nous avons déterminé un certain nombre de secteurs qui se prêtent à l'automatisation et nous avons déjà commencé à tester certains d'entre eux.

L'un de ces secteurs est la géographie. Les documents géographiques sont indispensables au succès d'un recensement pour deux raisons: d'abord des cartes exactes et lisibles aident les recenseurs à trouver tous les logements pour assurer un dénombrement complet; ensuite, des limites exactes et des données géographiques facilitent l'assignation de chaque logement et ses occupants à la superficie appropriée. Un des problèmes que nous avons connu au recensement de 1980 était que nos documents géographiques, y compris les cartes, étaient le produit d'opérations distinctes et avaient donc nécessité un volume considérable de travail d'écritures. Cette façon de procéder, lente et sujette aux erreurs, a entraîné des retards dans la production ainsi que des erreurs et des incohérences dans certains des résultats.

En vue du recensement de 1990, nous automatisons notre système de soutien géographique que nous appelons TIGER (Topologically Integrated Geographic Encoding and Referencing system). Le TIGER regroupera en un seul fichier toute l'information géographique qui était produite en plusieurs étapes distinctes en 1980. Il nous permettra de fournir en 1990 des produits et des services géographiques à partir d'une base de données homogènes et nous aidera à éviter certains retards et certaines inexactitudes du recensement de 1980. La génération par ordinateur de cartes qui correspondent aux régions géographiques de nos totalisations sera une grande amélioration par rapport aux opérations manuelles du recensement de 1980 et des recensements antérieurs. Un autre document qui sera présenté à cette conférence contient une description plus détaillée du système géographique automatisé.

Une autre amélioration prévue pour le recensement de 1990 est la création d'un fichier automatisé de contrôle des adresses. En 1980, même si la liste initiale de contrôle des adresses était informatisée, les changements apportés à la liste durant le recensement l'ont été manuellement. En 1990, nous aurons accès en permanence au fichier automatisé de contrôle des adresses de sorte que nous pourrons tenir la liste à jour. Nous avons déjà implanté avec succès un fichier automatisé de contrôle des adresses lors de nos recensements d'essai de 1985 et nous allons mener d'autres tests.

Avec un fichier automatisé d'adresses, il sera beaucoup plus facile de déterminer si une adresse spécifique figure ou non au fichier. Il sera aussi possible d'ajouter au fichier une adresse oubliée précédemment. Nous pouvons utiliser la technologie du code à barres pour informatiser le contrôle à l'arrivée des questionnaires. En conséquence, le personnel affecté au recensement pourra plus facilement déterminer les adresses pour lesquelles aucun questionnaire n'a été renvoyé et nous pourrions envoyer des lettres de rappel à ces adresses de manière à réduire encore plus le nombre de ménages non répondants auxquels doivent rendre visite les recenseurs. Enfin, nous pouvons mettre à jour une liste automatisée d'adresses et nous en servir dans les

opérations ultérieures du Census Bureau. Aux recensements d'essai de 1985, nous avons introduit avec succès un fichier automatisé de contrôle d'adresses, une fonction automatisée de contrôle à l'arrivée et l'utilisation de fiches de rappel.

L'une des façons les plus prometteuses de tirer profit de l'automatisation du recensement, et notre plus gros défi, consiste à convertir les données des questionnaires en un format lisible par ordinateur plus tôt dans le recensement que par le passé. Cette démarche est essentielle si nous voulons mettre à profit l'automatisation et diffuser plus rapidement les résultats. En 1980, la conversion des données n'a débuté qu'après la fermeture des bureaux de recensement provisoires et l'expédition des questionnaires à l'un des trois centres de traitement informatique. Au recensement de 1990, nous voulons commencer la conversion des données en même temps que la collecte. Ce début anticipé (de 5 à 7 mois avant la date au calendrier du recensement de 1980) laissera plus de temps à la revue et aux corrections, et permettra d'utiliser l'ordinateur pour certaines opérations du recensement. Il assurera un contrôle plus serré des affectations de suivi sur place et permettra de déceler plus tôt les problèmes de dénombrement. De même, les enregistrements informatiques des questionnaires pourraient servir de copies de sauvegarde si jamais les originaux étaient accidentellement détruits.

Bien que la décision d'implanter plus tôt le traitement informatique lors du recensement de 1990 fasse l'unanimité, deux questions importantes sont toujours sans réponse. Où le traitement automatisé sera-t-il exécuté et quelle technologie sera utilisée pour convertir les données des questionnaires en un format assimilable par ordinateur?

Pour ce qui a trait à la première question, il est utile de considérer deux grands scénarios pour l'accomplissement de cette conversion anticipée des données. Selon l'un de ces scénarios, les bureaux de district et de traitement seraient combinés et s'occuperaient à la fois du traitement automatisé et du suivi sur place. Il est fort peu probable que nous utilisions cette formule de bureaux "combinés" dans tout le pays à cause des difficultés que présentent la création, l'installation, l'intégration et la surveillance de 500 systèmes distincts de traitement des données; cependant, cette formule peut être adoptée dans les régions davantage rurales. Nous mettrons à l'épreuve un bureau "combiné" au cours de notre recensement d'essai de 1986 au Mississippi.

Suivant l'autre scénario, nous aurions des bureaux de district et de traitement distincts. Les bureaux de traitement recevraient les questionnaires envoyés par la poste par les répondants, les contrôleraient automatiquement à l'arrivée, convertiraient les données en un format lisible par machine et exécuteraient le contrôle automatisé des questionnaires. Les bureaux de district seraient uniquement chargés de contacter les ménages dont le questionnaire serait manquant ou incomplet. Nous avons testé ce plan lors de nos recensements d'essai de 1985 dans les bureaux de collecte de Jersey City au New Jersey et de Tampa en Floride, et le traitement à notre bureau permanent de traitement à Jeffersonville en Indiana, et tout s'est très bien déroulé. Au recensement d'essai de 1986 dans le comté de Los Angeles, nous aurons aussi recours à des bureaux de district et de traitement distincts, mais les deux seront situés dans la même région métropolitaine. Il est peu probable que nous utilisions des bureaux "distincts" dans tout le pays à cause des problèmes de communications et de soutien que cela poserait si les bureaux de traitement étaient éloignés des bureaux de district.

La formule des bureaux de traitement et de district combinés dans certaines parties du pays à faible densité de population et des bureaux de traitement et de district distincts dans les régions davantage urbanisées sera probablement l'option retenue pour le recensement de 1990.

Non seulement devons-nous décider où convertir les données en un format assimilable par machine en 1990, mais nous devons également déterminer comment nous allons procéder. En 1980, après avoir terminé les activités de collecte, nous avons introduit les données des questionnaires sur des bandes pour ordinateur en microfilmant les questionnaires (après le codage à la main des réponses en toutes lettres), puis en lisant les microfilms au moyen de lecteurs optiques (FOSDIC - Appareil de lecture optique pour transfert informatique). Pour le recensement de 1990, nous avons essentiellement le choix entre trois technologies ou toute combinaison de ces dernières. Nous pouvons continuer de procéder par transfert de microfilm à bande comme en 1980, mais avec du matériel plus récent et meilleur. Nous pouvons essayer de supprimer l'étape du microfilmage et de lire les questionnaires directement comme le sont les tests d'aptitudes dans les universités au moyen de la technologie de reconnaissance optique des caractères. Enfin, nous pouvons introduire les données au clavier. Il est peu probable que la conversion des données se fasse par introduction au clavier dans tous les centres de traitement, mais nous devrons utiliser beaucoup cette dernière méthode pour saisir dans l'ordinateur les adresses et les réponses en toutes lettres sur les questionnaires.

Aux recensements d'essai de 1985, nous avons employé les méthodes de reconnaissance optique des marques et d'introduction au clavier. Pour ceux de 1986, nous aurons recours aux méthodes d'introduction au clavier et de transfert de microfilm à bande. Malgré quelques problèmes, le lecteur optique de marques a suffisamment fait ses preuves pour que nous envisagions la possibilité d'en poursuivre l'essai. L'un des problèmes rencontrés était que le lecteur optique de marques disponible à l'époque ne pouvait traiter que du papier de $8\frac{1}{2}$ po sur 11 po. Il nous a donc fallu faire entrer tout le questionnaire abrégé sur une petite page, ce qui a limité la conception du questionnaire et pourrait expliquer les réponses inexactes faites à plusieurs questions. Si un lecteur pouvant accepter des feuilles de 11 po sur 17 po peut être mis au point, nous procéderons à un essai "sous contrôle" l'année prochaine.

La question des méthodes de conversion des données est liée aux formules de bureaux présentées ci- dessus, mais n'en dépend pas. Toute décision concernant le matériel fait aussi intervenir de nombreuses autres considérations, notamment le contenu et l'aspect des questionnaires et la facilité avec laquelle les répondants peuvent les remplir, la fiabilité et la disponibilité du matériel, les besoins en personnel déterminés par le matériel, c.-à-d. le nombre et les spécialisations des employés nécessaires ainsi que le coût et la maintenabilité du matériel.

Un autre document qui sera exposé à cette conférence abordera plus en détail le traitement en simultané (ou décentralisé). Nous allons augmenter ou améliorer l'automatisation dans d'autres secteurs pour accélérer le recensement et le rendre plus précis, et je parlerai brièvement de quelques-uns de ces secteurs.

L'un de ces secteurs est le contrôle des questionnaires. Il s'agit d'une tâche répétitive et monotone qui convient mieux à l'ordinateur qu'à l'être humain. La saisie des données des questionnaires plus tôt dans le recensement permettra de soumettre les données à un contrôle par ordinateur plus tôt que jamais. Ces contrôles visent à vérifier l'intégralité et la cohérence des données. En 1980, les questionnaires étaient contrôlés manuellement dans les bureaux de district essentiellement pour s'assurer qu'ils avaient été remplis en entier; puis, lorsque les questionnaires étaient lus par les lecteurs FOSDIC, l'ordinateur les soumettait à un contrôle pour s'assurer qu'ils étaient complets et cohérents. En 1990, la plupart des contrôles manuels seraient éliminés, assurant ainsi un contrôle plus rapide, plus cohérent et plus précis. La décentralisation de l'opération de contrôle par ordinateur nous permettrait, au besoin, de communiquer de nouveau avec le répondant, ce qui était impossible au recensement de 1980.

Une autre technique d'automatisation prometteuse concerne le codage des réponses en toutes lettres sur le questionnaire. En 1980, le codage manuel des réponses en toutes lettres a été considérable, long et coûteux. En 1990, nous allons introduire au clavier les réponses en toutes lettres et un logiciel spécial attribuera à ces réponses les codes lisibles par ordinateur correspondants. Nous ne pouvons pas supprimer complètement le travail d'écritures du codage, étant donné que certaines réponses en toutes lettres seront incomplètes ou impossibles à coder et devront être soumises à nos services de renvois. Nous serons cependant en mesure de réduire considérablement le volume de travail manuel et donc d'économiser du temps et de l'argent, et d'améliorer la qualité des données. Nous prévoyons tester certains aspects du codage automatisé lors de nos recensements d'essai en 1986.

Nous ferons aussi appel à l'automatisation pour mieux planifier et surveiller le recensement. Le Census Bureau est en train de créer une base de données complexe d'information de la gestion pour faire en sorte que nous respections les principales échéances dans la prise de décisions au sujet de la forme que prendra le recensement de 1990. Le système d'information de la gestion était en place pour nous aider à suivre les opérations des recensements d'essai de 1985 et à planifier ceux de 1986. Non seulement ce système d'information de la gestion nous aidera-t-il à planifier le recensement de 1990, mais il nous fournira des données à jour sur les coûts et sur l'état d'avancement des travaux qui nous permettront de surveiller le déroulement proprement dit des opérations du recensement de 1990. En 1980, les états de coûts et les rapports sur l'état d'avancement n'étaient pas intégrés à d'autres rapports de gestion, et certaines des données sur les coûts et sur l'état d'avancement des travaux mettaient plusieurs jours à parvenir aux gestionnaires.

L'automatisation nous aidera à contrôler et à surveiller bon nombre d'autres fonctions administratives. Nous aurons un système automatisé de paye comme en 1980. Pour le recensement de 1990, nous aurons également, sur micro-ordinateur, un nouveau fichier d'employés qui nous permettra de mieux organiser l'information voulue sur notre important effectif temporaire. (Nous avons fait cela aux recensements d'essai de 1985.) Par exemple, nous saurons si nous respectons ou non les objectifs de recrutement que nous nous sommes fixés dans chaque secteur de dénombrement et nous pourrons utiliser le fichier pour constituer les tâches des recenseurs. Nous

aurons également un nouveau système automatisé de gestion des stocks pour gérer l'achat et la distribution de la grande quantité de fournitures spécialisées nécessaires à la tenue du recensement.

Enfin, nous envisageons d'automatiser les opérations de totalisation et de publication du recensement de 1990. Notre système de totalisation a été complètement informatisé pour le recensement de 1980, mais, pour celui de 1990, nous comptons améliorer le système en mettant à profit les progrès en logiciel de base de données. Nous voulons aussi utiliser l'ordinateur dans l'examen analytique des données totalisées, fait manuellement en 1980. Cet examen qui consiste à chercher les erreurs et les anomalies dans les données est indispensable pour maintenir la qualité de nos résultats. L'ordinateur servira à accélérer et à améliorer cette analyse.

De nouvelles techniques d'automatisation joueront également un rôle dans la diffusion des résultats du recensement de 1990. Le Census Bureau continuera de produire des rapports imprimés et de volumineux fichiers sommaires sur bande pour ordinateur, mais doit également répondre aux besoins des utilisateurs de petits ordinateurs qui veulent des sorties sur disquette. Une autre innovation que nous envisagerons pour 1990 sera une base de données sommaires que les utilisateurs pourront consulter en direct à partir de leurs ordinateurs de bureau par liaison téléphonique. Le Census Bureau a déjà implanté à une échelle limitée un tel système appelé CENDATA. Il y aura peut-être d'autres changements dans les quelques années à venir, notamment des améliorations aux disques laser, que nous pourrons mettre à profit dans le recensement de 1990. Heureusement, les décisions finales au sujet des totalisations et des résultats peuvent être prises plus tard au cours de la décennie, nous permettant de tirer ainsi profit de nouvelles technologies.

Malgré la surexcitation que provoquent au Census Bureau ces possibilités d'automatisation, il convient de faire quelques mises en garde. En effet, quels que soient les systèmes élaborés, ils doivent être simples, étant donné qu'ils seront exploités par un effectif temporaire dont la formation sera minimale. Les systèmes doivent être entièrement testés, s'avérer fiables et être essentiellement "à sécurité intégrée" pour éviter les pannes paralysantes. Le matériel ne doit pas être exagérément coûteux et devrait continuer d'être utile au Census Bureau ou être vendable après le recensement.

Ce qui importe surtout dans notre quête d'automatisation du recensement, c'est de faire en sorte d'assurer, en fait et en apparence, la confidentialité des données que nous recueillons. La garantie de confidentialité du processus de recensement est la seule façon d'obtenir la confiance et la collaboration du public. Le Census Bureau est fier de sa tradition de confidentialité et est à la recherche constante de moyens de garantir et d'améliorer cette confidentialité. Le Census Bureau ne divulgue ses données sur les particuliers à personne, pas même à d'autres organismes fédéraux. Toutefois, les risques de la technologie nous obligent à déployer plus d'efforts pour convaincre les gens que le fait de répondre au questionnaire du recensement ne leur est aucunement préjudiciable et que les renseignements qu'ils nous fournissent sont strictement confidentiels aux termes de la loi.

Conclusion

Nous devons prendre des décisions au sujet de ces deux grandes questions (où et comment) se rapportant à la conversion des données d'ici septembre 1986 pour pouvoir amorcer la phase d'acquisition du matériel. Certains ont proposé que nous prenions ces décisions plus tôt et nous allons accélérer le processus décisionnel partout où nous le pourrons. Nous croyons cependant qu'il est important de tirer le plus possible parti des résultats de nos recensements d'essai avant de prendre des décisions aussi graves. Nous organisons à la mi-octobre 1985 une conférence sur le processus décisionnel du recensement décennal (Decennial Census Decision Conference) où nous traiterons plusieurs questions, notamment "où" se fera la conversion.

Bien qu'il reste beaucoup de décisions à prendre et de problèmes à résoudre, nous sommes assez avancés dans notre processus de planification de l'automatisation pour affirmer que le recensement de 1990 sera considérablement plus automatisé que tout autre recensement antérieur. Nous ferons une utilisation innovatrice de techniques d'automatisation pour procéder à la saisie des données plus tôt que jamais. Nous aurons un système automatisé de soutien géographique. Nous contrôlerons les questionnaires par ordinateur. Enfin, nous avons déjà implanté un fichier automatisé de contrôle des adresses, une fonction automatique de contrôle à l'arrivée des questionnaires et un système automatisé d'information de la gestion dans nos recensements d'essai de 1985 et nous prévoyons les conserver en 1990.

OPTIONS EN MÉTHODES DE SAISIE DES DONNÉES

DAVE A. CROOT

DIVISION DES SERVICES AUX CLIENTS
STATISTIQUE CANADA

Introduction

Depuis l'avènement de l'ordinateur, la puissance de traitement a augmenté de façon spectaculaire. La recherche parallèle de méthodes de saisie des données efficaces et pratiques pour alimenter ces machines voraces a amené des innovations également à un rythme accéléré, sinon avec le même effet dramatique.

Les méthodes de saisie des données peuvent être classées en deux catégories: celles qui font intervenir un opérateur humain et celles qui dépendent (presque) entièrement des machines. À l'origine, la saisie de toutes les données "de gestion" se faisait par des opérateurs sur des machines à clavier et à mécanisme de perforation. Le support de sortie était généralement une carte de 80 colonnes ou une bande de papier perforée. Malgré le progrès technologique que ces machines ont connu avec le temps et l'augmentation correspondante de leur rendement, la présence d'un opérateur humain reste toujours la caractéristique principale de cette classe de matériel de saisie des données.

Certaines des premières applications scientifiques des ordinateurs réalisaient la saisie directe des données au moyen d'une forme quelconque de capteur relié à des instruments de laboratoire ou de processus industriel qui mesuraient des phénomènes physiques comme la température ou le débit. Ces caractéristiques ont été améliorées avec le temps et la lecture des données par machine s'est étendue au monde des affaires, c'est-à-dire aux chiffres et aux mots inscrits sur papier.

La saisie des données, par l'être humain ou par machine, comprend trois étapes:

1. Reconnaissance
2. Interprétation
3. Enregistrement

Saisie des données par l'être humain

Dans les méthodes à intervention humaine, la carte ou la bande perforée que la machine lisait les premiers temps a cédé le pas à l'enregistrement des données sur des supports magnétiques beaucoup plus rapides et plus pratiques. Le perfectionnement des systèmes de communication permet la transmission directe à des centres éloignés, éliminant le besoin de mémoires auxiliaires régionales où les données sont recueillies en attendant leur traitement par l'ordinateur central. Toutefois, la plupart des progrès technologiques intéressent la première étape du processus. Avec la venue sur le marché d'ordinateurs moins chers et plus facilement accessibles, il a été possible de regrouper la saisie des données et d'autres opérations connexes comme le contrôle et le codage. En évitant la manipulation répétée des mêmes documents, on réalise des économies directes en ressources humaines mais, ce qui est peut-être plus important, on peut aussi réaliser des gains de précision. Un exemple particulièrement impressionnant de cette sorte de progrès est l'ITAO (interview téléphonique assistée par ordinateur) où la communication avec le répondant, la collecte et la saisie des données et un certain contrôle sont tous exécutés en une seule étape. Ces augmentations de la puissance de calcul achetable au coût unitaire se sont accompagnées de gains similaires dans la puissance de calcul accessible pour un volume et un poids donnés. On peut aujourd'hui obtenir un ordinateur d'une puissance considérable avec une mémoire à accès direct assez grande (256K) et des fonctions de programmation raisonnables dans une unité de la taille et du poids d'un téléphone portatif. Cette accessibilité à la puissance de calcul donne à des terminaux de saisie de données pesant une livre au maximum la capacité de saisir, de contrôler et de mémoriser les données de plusieurs centaines d'enregistrements d'enquêtes. Ainsi, on pourrait avoir une opération de saisie des données mobile qui fonctionnerait plusieurs jours avant de devoir télécharger les données dans un ordinateur central.

En général, on peut affirmer que tous les progrès réalisés en saisie des données à intervention humaine sont parties de la reconnaissance d'une activité exécutée au rythme de l'homme et ont essayé de rendre cette activité aussi complète et exhaustive que possible sans en réduire le rythme.

Saisie automatique des données

Le premier exemple de saisie automatique des données appliquée est celui des appareils de reconnaissance magnétique de caractères utilisés par les grands établissements bancaires. Le but initial était la lecture et le tri rapide des chèques dans les chambres de compensation. Les mêmes

caractères employés alors se retrouvent aujourd'hui au bas des chèques, mais l'encre magnétique n'est plus utilisée. La lecture des chèques est aujourd'hui optique et non plus magnétique. La deuxième réalisation a été la lecture de caractères optiques hautement stylisés que l'on retrouve sur les chèques et qui, comme nous l'avons dit, ressemblent de près aux caractères optiques qu'ils ont remplacés. Ici aussi, ce sont les banques ou les organismes à besoins similaires qui ont été les premiers à utiliser la reconnaissance optique de caractères pour traiter de très grands nombres de transactions comparativement simples. Ces premières inventions étaient caractérisées par des jeux de caractères spécialisés limités, à distinction très nette entre les caractères pour permettre leur reconnaissance franche par le mécanisme de lecture. Avec une résolution limitée, il s'agissait avant tout de faciliter l'interprétation de la machine, généralement aux dépens de la convivialité. Dans la plupart des cas, les caractères étaient d'abord imprimés par une autre machine, ce qui permettait de garder le contrôle du principal facteur de performance de la lecture de caractères qui devait suivre. Il y a eu une amélioration constante de la vitesse et de la résolution de la lecture, ce qui a permis une meilleure humanisation des jeux de caractères qui pouvaient être lus automatiquement. La lecture de caractères s'appuie sur une des deux techniques ci-dessous:

1. la représentation d'une image numérique du caractère objet en une matrice de points comparée ensuite à des matrices de caractères spécifiées, ou

2. l'utilisation du tracé du caractère pour la comparaison.

Quelle que soit la technique utilisée, il est certain que l'amélioration de la résolution rend l'interprétation plus exacte. Les premiers caractères magnétiques et leurs cousins optiques étaient très stylisés pour en faciliter l'interprétation même avec une résolution limitée. L'avancement de la technologie permet aujourd'hui de lire avec un degré élevé de précision des caractères dactylographiés de types et de tailles multiples (tels que l'on serait susceptible de recevoir d'un nombre de répondants différents).

Une autre grande réalisation a été, au début des années 1950, la reconnaissance de données écrites, initialement de marques seulement plutôt que de caractères alphanumériques. C'est ainsi que le lecteur FOSDIC (Appareil de lecture optique pour tranfert informatique) a été inventé par le U.S. Bureau of the Census et utilisé au Canada pour les recensements de la population et du logement de 1971 et de 1976. Avec ce système, les questionnaires sont d'abord filmés. La lecture en est donc comparativement positive, les marques étant reconnues à partir d'une lumière transmise et non réfléchie. La reconnaissance est d'autant plus facilitée que seule la présence ou l'absence de marques dans des positions données doit être déterminée. La signification est déduite des positions qu'occupent les marques. De nombreuses questions du recensement ou d'autres enquêtes sociales ont un petit ensemble de réponses possibles et se prêtent donc naturellement au format à choix multiple et, par conséquent, à la représentation de la réponse par des marques optiques. Mais, étant donné qu'il est rarement possible, excepté avec des questionnaires extrêmement structurés, de traiter toutes les questions de cette manière, la saisie des caractères ou des chiffres s'impose. Il y a aujourd'hui des machines qui lisent partiellement les caractères manuscrits. Compte tenu des caprices de l'impression manuelle, sans parler de l'écriture manuelle, la précision d'une telle lecture reste suspecte et généralement le système exige une certaine intervention humaine. Lorsqu'un caractère est lu, il est numérisé et la matrice ou la représentation numérique de l'image qui résulte de la lecture est ensuite assortie par le système à une série d'entrées de son fichier de caractères potentiels. Les systèmes plus puissants ont de plus gros fichiers de caractères potentiels et peuvent reconnaître différentes représentations d'un même caractère. Ils peuvent aussi "apprendre" d'autres formes particulières en ajoutant de nouvelles entrées aux fichiers à partir du texte lu. L'intervention humaine consiste à rendre une décision lorsque le système ne peut assortir le caractère lu à aucun des caractères enregistrés dans ses fichiers. Le système soumet alors la représentation numérique qu'il a gardée à l'interprétation humaine. Pour cela, il affiche le contexte, c'est-à-dire les caractères reconnus dans chaque enregistrement contenant un ou plusieurs caractères non identifiés aux images numériques d'un caractère quelconque non identifié. L'intervenant humain, dont l'agilité mentale est plus grande que celle de la machine, peut alors attribuer un sens à l'image. Ces interprétations peuvent être utilisées de façon heuristique, lorsque les variantes des caractères sont susceptibles d'être répétitives ou simplement comme une étape du traitement opérationnel.

Une autre voie explorée de la saisie automatique des données a été la reconnaissance automatique de la parole. La description ci-dessus de l'évolution de la lecture optique de caractères pourrait être utilisée avec quelques modifications pour la reconnaissance de la parole. Après un départ tardif, l'évolution a été rapide bien que le processus

soit intrinsèquement complexe. Cette complexité vient de la dimension additionnelle de la variabilité causée par les différences entre les voix féminines et masculines, par les différences d'accent et de prononciation des voyelles dans les différents parlers nationaux et régionaux, et du fait que le discours normal est, encore plus que l'écriture, composé d'une chaîne continue plutôt que de caractères nettement séparés. En ce moment, des vocabulaires de 500 à 1,000 mots peuvent être traités avec une précision raisonnable. Comme cette méthode s'emploie surtout en temps réel, le système peut nous donner l'occasion de reformuler plus lentement des mots qu'il n'aurait pas reconnus une première fois. L'utilisation de la technique semble graviter autour de certaines applications spécialisées où l'accès direct aux fichiers en mémoire vient compenser une activité relativement lente, comme le signalement de symptômes dans un diagnostic médical ou l'assistance annuaire téléphonique.

Le choix pour 1991

Le choix essentiel pour la saisie des données du recensement de 1991 se situe entre l'introduction et la saisie automatique par machine, avec dans le dernier cas presque certainement une forme de lecture optique de caractères ou de marques, ou une combinaison de ces solutions. Bien que des progrès impressionnants aient été réalisés dans ce domaine, nous restons confinés aux caractères séparés prévisibles, loin de la vraie écriture manuscrite. Dans une opération comme le recensement, il serait extrêmement avantageux de saisir un très grand pourcentage des données sans intervention humaine par la suite. Pour cela, il faudrait construire de puissants algorithmes de reconnaissance des formes et disposer d'une très grande puissance de calcul. L'industrie semble sur le point d'avoir des algorithmes adéquats. Cependant, la puissance de calcul exigée, avec le matériel optique haut de gamme et l'équipement mécanique complexe de manipulation du papier nécessaires requièrent des investissements considérables en matériels.

Plus les machines utilisées sont automatisées, plus elles coûteront cher. Le coût est un facteur critique pour une opération comme le recensement qui se caractérise par un très grand volume de données, des délais serrés et une fréquence rare.

Il est certes possible de démarrer et d'arrêter au moins la partie travail d'une activité à intensité-travail de courte durée comme la saisie des données du recensement. En fait, la plupart des autres opérations du recensement qui précèdent l'étape du contrôle automatique sont traitées exactement de cette façon. Par ailleurs, le coût du travail reste généralement constant lorsque les délais sont rapprochés. Avec une méthode à intensité capitalistique, le rapprochement des délais nécessiterait probablement plus d'équipement et augmenterait presque certainement le coût du traitement.

Nous avons été particulièrement heureux en 1981 et en 1986 d'avoir pour le recensement de la population un système d'introduction par clavier complet et très efficace, qui a été construit sur l'infrastructure que gardait Revenu Canada (Impôt) pour son traitement annuel. Une étude comparative faite avant le recensement de 1981 laissait prévoir une légère économie avec ce système d'introduction par rapport au système FOSDIC qui avait été utilisé pour les deux recensements précédents. Une deuxième économie a été réalisée avec le recours pour cette opération à Revenu Canada (Impôt) qui avait la plupart du matériel et, comme nous l'avons déjà dit, toute l'infrastructure nécessaire. Dans tout le processus, le coût de l'introduction, de la vérification et de la transmission des centres régionaux à Ottawa a été de moins de 50 cents par ménage et le temps écoulé a été de deux à trois mois plus court qu'avec le système (FOSDIC) de 1976.

Il y a de nouvelles tendances intéressantes en lecture optique vers des systèmes moins performants mais moins chers que les gros lecteurs optiques du haut de gamme. L'échelle de ces appareils convient au poste de travail à un opérateur attribuant les caractères non reconnus. Ce type de machine serait toujours utilisée pour la saisie des données dans les régions et pourrait être un compromis raisonnable dans le degré d'automatisation à l'avenir. Une autre innovation intéressante est le stockage des données sur disque optique. FOSDIC, nous l'avons déjà dit, nécessite le filmage préalable des questionnaires, mais comme le recensement doit garder une image permanente du questionnaire, cette étape sert un double objet. Avec la construction de lecteurs à haute performance, l'image numérique de chaque questionnaire pourrait être également gardée comme image permanente du questionnaire, en plus de son rôle dans la saisie des données. Plusieurs systèmes commerciaux de lecture optique dotés de la capacité de mémoire inaltérable massive des disques optiques et de terminaux d'affichage ou d'imprimantes sont en train de paraître sur le marché. Avec les lecteurs-interpréteurs de bas de gamme mentionnés plus haut, ces nouveaux systèmes pourraient permettre une combinaison de lecture automatique et d'introduction au clavier à haut rendement.

Conclusion

Aux prix actuels, la saisie par lecture automatique ne peut concurrencer l'introduction au clavier si le matériel n'était utilisé que pour le recensement. Des progrès sont réalisés et il y a une forte probabilité d'une faisabilité technique avant 1991. Toutefois, il ne sera probablement pas économique d'utiliser une méthode automatique de saisie des données à intensité capitalistique pour le recensement à moins que le même matériel ne soit utilisé pour un grand nombre d'autres enquêtes afin d'étaler le coût en capital sur plusieurs opérations permanentes.

DÉCENTRALISATION DE LA SAISIE DES DONNÉES DU RECENSEMENT DE 1990

ARNOLD A. JACKSON

DECENNIAL OPERATIONS DIVISION
U.S. BUREAU OF THE CENSUS

Introduction

Le Bureau of the Census s'est donné un objectif ambitieux pour le recensement de la population et du logement de 1990, soit celui de faire le recensement décennal suivant à un coût par logement égal ou inférieur au coût du recensement de 1980 (en dollars constants). Au même moment, les pressions exercées sur le Census Bureau pour qu'il produise plus tôt un plus grand volume de données de meilleure qualité ne cessent de croître. Inébranlés par la diminution des ressources fédérales réelles, les utilisateurs des données du Census Bureau exercent des pressions de tous les côtés et s'attendent à plus pour moins.

Il n'est pas étonnant qu'avec des exigences difficilement conciliables à satisfaire, à savoir réduire les coûts tout en produisant autant ou plus qu'avant, le Census Bureau cherche dans l'automatisation une solution éventuelle. Même si le recensement de 1990 ne sera pas complètement automatisé, toutes les étapes importantes du recensement subissent un examen approfondi qui sera suivi d'une étude détaillée et d'essais sur le terrain pour confirmer les possibilités d'automatisation appropriées pour 1990.

Ces étapes sont les suivantes:

- réception et contrôle à l'arrivée des questionnaires renvoyés par la poste par les répondants et remplis par les recenseurs (plus de 100,000,000 sont prévus en 1990);
- vérification des adresses des ménages qui ont renvoyé des questionnaires par îlot (contrôle des adresses);
- contrôle du contenu et de la couverture;
- conversion du contenu acceptable en un format lisible par machine; et
- production de totalisations pour la publication de résultats du recensement.

En 1990, l'exécution de ces étapes nécessiterait plus de 60,000 commis à un coût de près de 1 milliard de dollars sans automatisation. Elles ont été réalisées dans **trois** centres en 1980 pour fournir des chiffres provisoires à l'administration centrale du Census Bureau. Ces opérations séquentielles ont duré sept mois, les données du recensement ayant été converties en un format lisible par machine seulement après le passage de toutes les formules d'un bureau de district par toutes les étapes antérieures selon un mode de traitement par lots.

Ce recensement coûteux, sujet aux erreurs et difficile à gérer a amené le Census Bureau à considérer sérieusement les avantages éventuels que représenterait le traitement simultané par rapport au traitement séquentiel. Le traitement simultané permettrait notamment:

- de débarrasser les centres de collecte et de traitement des formules originales à mesure qu'ils ont fini de les traiter;
- de déterminer les questionnaires dont le contenu ou la couverture pose des problèmes aux étapes initiales de la collecte pour accélérer le travail de suivi et faire en sorte que les recherches, appariements et autres raffinements de couverture se fassent le plus rapidement possible. Les suivis doivent débuter pendant que les bureaux de district comptent encore du personnel compétent sur le terrain;
- de cerner et de résoudre les problèmes de collecte et de traitement propres à certaines régions géographiques à temps pour pouvoir trouver et prendre des mesures correctrices spéciales;
- de terminer toutes les opérations de collecte, de traitement et de suivi sur place, et les revues locales à temps pour produire les totalisations finales (avec les rajustements nécessaires) bien avant la date d'utilisation des chiffres du recensement pour la répartition des sièges de la U.S. House of Representatives (le 31 décembre 1990).

En conséquence, le Census Bureau croit que les options qui lui permettraient de suivre cette méthode de collecte et de traitement sont celles qui sont décentralisées. La décentralisation, au sens où elle est employée dans le reste du document, consiste à exécuter les opérations de collecte

et de traitement jusqu'à l'étape des totalisations finales dans un grand (20 et plus) ou petit (de 5 à 20) nombre de centres éloignés de l'administration centrale du Census Bureau.

Options fondamentales

Une foule d'options de décentralisation permettraient au Census Bureau d'exécuter simultanément la collecte et le traitement des données. Un certain nombre de possibilités sont actuellement envisagées. Toutes impliquent des frais que le Census Bureau doit assumer pour le financement du matériel, du logiciel, des locaux et du soutien administratif nécessités par la collecte et le traitement simultanés des données. Ainsi, la véritable question qui se pose est la suivante: "les avantages de l'exécution simultanée des opérations de collecte et de traitement justifient-ils le coût de la décentralisation qu'elle suppose?". Nous croyons fermement que le coût de l'exécution simultanée de ces opérations dans un contexte centralisé est prohibitif en raison des services de soutien et de l'ampleur du type d'installations qui seraient nécessaires.

Ce document traite deux des nombreuses options à l'étude. Il vise à provoquer aujourd'hui une discussion sur les besoins, sur les avantages et les inconvénients et sur les répercussions stratégiques des choix qui s'offrent à nous. Les deux options exposées ici sont les suivantes:

1. Décentralisation intégrale
   (Pièce A)

2. Décentralisation régionale
   (Pièce B)

Deux raisons ont motivé le choix de ces options. Premièrement, l'option de la décentralisation intégrale est la plus extrême des options examinées. Elle supposerait la mise sur pied de 350 à 450 bureaux, contrairement aux autres options où les centres sont considérablement moins nombreux. Deuxièmement, elles sont toutes deux fondées sur le réseau national de bureaux de collecte de district et de bureaux régionaux des opérations sur le terrain de notre recensement de 1980. Cette année-là, les bureaux de district ont trié, réceptionné, contrôlé et autrement préparé les formules pour la saisie des données qui s'est déroulée dans l'un des trois seuls centres aménagés à cette fin. À la structure permanente des bureaux régionaux d'opérations sur le terrain ont été ajoutés des centres de recensement chargés de gérer les bureaux de district desservant chacun quelque 250,000 ménages. En outre, les avantages éventuels et les coûts imposants sont plus clairs dans ces deux options extrêmes fort simples que dans les options hybrides. Le texte qui suit donne une description détaillée des deux options, suivie des avantages et inconvénients de chacune.

La dernière section du document traite certaines répercussions stratégiques qui relient notre analyse de chacune des options de collecte et de traitement aux objectifs que nous nous sommes fixés pour le recensement de 1990 relativement aux échéances, à la qualité, au coût, à la collaboration de la population et à la réduction des risques.

Décentralisation intégrale

Selon cette option, les activités de collecte et de traitement sont exécutées sous un même toit dans les bureaux de district. Ces bureaux seraient responsables du contrôle à l'arrivée des questionnaires, de la mise à jour du fichier des adresses, du contrôle des questionnaires, de la conversion des données, des recherches et des appariements, et des opérations de suivi nécessaires pour assurer la qualité des données, notamment des visites aux ménages pour corriger les formules-échantillon rejetées au contrôle.

Chaque bureau de district étant un centre de recensement pratiquement autonome, aucun document n'est envoyé d'un bureau à l'autre. La conversion des données se fait directement au terminal ou par lecteurs de table. Les services de soutien sont relativement simples et la responsabilité comptable des gestionnaires peut être cernée clairement.

Le soutien des opérations sur le terrain serait à son niveau le plus élevé suivant cette option, étant donné que les affectations de suivi sur place seraient constituées à mesure que le fichier de contrôle des adresses produirait les états quotidiens d'avancement des travaux par îlot. Il est inutile de prévoir du temps pour la transmission des données dans cette option. Par ailleurs, le Census Bureau pourrait exercer son contrôle de gestion des opérations par l'entremise de la structure régionale en place. En même temps, il entreprendrait de relever le défi qu'est la gestion de 350 à 450 centres informatiques aménagés dans des locaux soigneusement choisis de 12,000 à 15,000 pieds carrés chacun, avec la sécurité, la climatisation et le câblage adéquats pour le traitement électronique des données. Chaque bureau de district aurait besoin d'un effectif d'environ 100 personnes, soit moins que les chiffres avancés dans les autres options envisagées à l'heure actuelle. Par ailleurs, le Census Bureau devrait ici trouver, former et garder à son emploi un nombre considérable de gestionnaires

capables de mener efficacement un bureau de collecte et de traitement combinés. Ce sont là quelques-unes des caractéristiques et des impératifs des opérations entièrement décentralisées de collecte et de traitement du recensement de 1990.

Avantages et inconvénients

Les aspects les plus attirants d'une décentralisation intégrale sont les suivants:

Dotation en personnel

- Nous n'aurions pas besoin d'un effectif aussi important à un même endroit et nous pourrions avoir de meilleurs employés, par exemple, en recrutant un pourcentage plus élevé d'opérateurs qui sont de bons dactylographes.

- Nos salaires seraient plus compétitifs dans les villes où ces bureaux seraient situés que dans les villes où seraient les bureaux plus centralisés. (Pour le recensement d'essai de 1986, nos salaires sont beaucoup plus compétitifs dans le East Central Mississippi que dans le comté de Central Los Angeles.)

Charge de travail

- Les bureaux de traitement seraient plus faciles à gérer et, par conséquent, moins sujets aux goulots d'étranglement et à l'asphyxie. Le principal problème de IRS est que ses nouveaux systèmes ne pouvaient pas s'acquitter des lourdes charges de travail.

Coordination du traitement et des opérations sur le terrain

- Le rapprochement du traitement de la collecte sur le terrain accélérerait les opérations sur le terrain. De fait, nous aurions plus de temps pour effectivement traiter les données et préparer les affectations sur le terrain.

Plans d'urgence

- En cas de sinistre, le travail serait distribué aux bureaux avoisinants le plus en mesure de l'absorber. Étant donné que tous les bureaux seraient petits, un incendie dans l'un d'eux qui détruirait des questionnaires aurait ici des conséquences beaucoup moins graves que dans tout autre scénario.

Entreposage des formules

- Comme les charges de travail sont petites, il est plus facile de classer les formules par région géographique (CBNA) pour faciliter, le cas échéant, leur consultation par les commis. En effet, ceux-ci auraient à consulter les formules si jamais nos routines de recherche et d'appariement connaissent des difficultés imprévues ou que, pour une raison quelconque, nous sommes incapables d'introduire des noms au clavier.

Les inconvénients correspondants consistent en obstacles éventuels dans les secteurs suivants:

Locaux

- Nous aurions besoin de locaux sûrs et climatisés dans plus de 400 bureaux. En effet, le matériel automatisé qui se trouvera dans les bureaux de district est quelque peu sensible à la chaleur et à l'humidité.

Personnel technique

- Bien que le degré de connaissances techniques nécessaires soit faible, l'embauchage de ce personnel peut s'avérer difficile à certains endroits.

Installation et entretien

- Depuis toujours, le Census Bureau installe et entretient son propre matériel. Il nous faudrait avoir recours à des entreprises privées pour installer et entretenir le matériel.

Contrôle de l'administration centrale

- L'administration centrale serait incapable d'exercer un contrôle aussi serré sur les opérations de traitement si celles-ci étaient décentralisées. La qualité serait toutefois assez uniforme puisque le logiciel déterminerait en grande partie la qualité et qu'aucun bureau à lui seul ne serait assez important pour réduire la qualité des données de tout le pays.

Mises à jour de logiciel

- Les mises à jour de logiciel devraient être transmises par téléphone et seraient plus difficiles à réaliser que dans un système de traitement plus centralisé.

Utilisation du matériel

- Une décentralisation intégrale nécessiterait plus de matériel de secours. En outre, nos prévisions des charges de travail devraient être meilleures que celles des recensements passés. Après l'installation du matériel dans les bureaux de district, il serait difficile de modifier leur configuration.

Régions à faible densité de population

- La dotation en personnel peut être difficile dans certains des bureaux de district à plus faible densité de population sans SMSA importante (régions primaires qui étaient courantes en 1980).

Décentralisation régionale

Selon cette option, le traitement serait séparé de la collecte. De cinq à vingt bureaux régionaux relèveraient des bureaux de collecte de district. Douze serait le nombre idéal suivant ce scénario, car ce chiffre correspond à celui de nos bureaux permanents d'opérations sur le terrain. Le nombre de bureaux de district servis par un bureau de traitement serait dans une large mesure déterminé par la densité de la population et par la capacité des gestionnaires de contrôler efficacement de 15 à 100 bureaux par l'entremise d'un bureau de traitement - même lorsqu'ils correspondent aux bureaux régionaux d'opérations sur le terrain.

Une décentralisation régionale limite le nombre de bureaux que l'administration centrale doit gérer et servir et nous permet d'utiliser plus de matériel automatique, mais au risque de rendre difficile la dotation en personnel des bureaux.

Selon ce scénario, le bureau de traitement reçoit tous les questionnaires renvoyés par la poste, introduit au clavier les noms des ménages non répondants devant faire l'objet de suivi, saisit les données, prépare le suivi pour les bureaux de district des questionnaires rejetés au contrôle, introduit les noms, fait des appariements pour améliorer la couverture et dresse les listes de suivis sur place à effectuer.

La conversion des données dans les bureaux régionaux de traitement serait faite au moyen des lecteurs FOSDIC du Census Bureau ou de lecteurs optiques de marques spécialement conçus à cette fin.

Les besoins en transmission concernent essentiellement le renvoi aux bureaux de district de 1.2 à 2.7 millions de questionnaires rejetés au contrôle en vue des suivis téléphoniques, de 33,000 à 75,000 pages de listes de suivis en cas de non-réponse et un nombre similaire de pages de listes de suivis en cas de rejet au contrôle. Le bureau de traitement recevrait régulièrement entre 3.5 à 8 millions de formules des bureaux de district (non-réponses, suivis dans les cas de rejet au contrôle et logements inoccupés et à supprimer).

Selon cette option, le soutien des opérations sur le terrain devrait être soigneusement coordonné étant donné que certains bureaux de district pourraient se trouver à une distance de 2 à 4 heures en voiture des bureaux de traitement. Étant donné que la majorité des données qui circulent entre ces points est assujettie aux dispositions du règlement 13 sur la confidentialité, il est peu probable que les télécommunications règlent ce problème.

Avantages et inconvénients

Les avantages éventuels d'une structure régionale sont les suivants:

Coût

- Les méthodes de lecture - en particulier la lecture des questionnaires complets et abrégés - sont les moins coûteuses. La lecture de toutes les données coûterait moins que d'autres options de conversion.

Contrôle de l'administration centrale

- L'option de décentralisation régionale faciliterait le contrôle par l'administration centrale compte tenu du petit nombre de bureaux.

Soutien technique

- Vu le petit nombre de bureaux, les techniciens du Census Bureau pourraient assurer l'entretien du matériel.

Uniformité opérationnelle

- Selon cette option, les centres régionaux de traitement seraient calibrés de manière à utiliser des matériels, des logiciels et des procédures d'exploitation similaires. Les techniques de gestion seraient interchangeables et le contrôle de la qualité serait plus facile à exercer que dans une structure hybride ou variée.

Cependant, certains inconvénients de cette option méritent également d'être signalés. Ce sont les suivants:

- Coordination du traitement et des activités sur le terrain

  Les gestionnaires des bureaux de traitement ne pourront peut-être pas établir des relations de travail avec le personnel sur le terrain en-deçà du niveau des bureaux régionaux. Les

systèmes de gestion du traitement et des opérations sur le terrain peuvent ne pas être parallèles, c.-à-d. que les bureaux de traitement peuvent dépasser les limites régionales et compliquer la coordination.

- Dotation en personnel

  Les plus gros bureaux de traitement seraient extrêmement difficiles à doter en personnel étant donné que leur taille serait de beaucoup supérieure à celle des bureaux de 1980.

- Technologie non rodée à cette fin

  Aucun lecteur optique de marques existant ne répond à nos besoins. Des machines sont annoncées, mais pas encore sur le marché. Aucun essai complet de l'utilisation du lecteur FOSDIC (Appareil de lecture optique pour transfert automatique) sans contrôle initial n'aura été fait avant la prise de décisions concernant la méthode de saisie des données. Le recensement d'essai de 1986 se fera à l'aide des appareils-photo de 1980 qui n'ont pas la résolution qui sera exigée des appareils de 1990.

Locaux

- La superficie d'un bureau moyen devrait être égale à celle de plusieurs terrains de football. Comme ces bureaux doivent avoir une source d'alimentation puissante et une excellente climatisation, nous aurions de la difficulté à trouver des locaux convenables.

Contrôle central

- Contrôler le flot de documents serait très difficile à cause du volume élevé de documents échangés à l'intérieur des bureaux de traitement et entre ceux-ci et les bureaux de district. Le risque serait élevé d'envoyer occasionnellement des documents au mauvais bureau de district, ce qui retarderait le travail sur le terrain de 2 à 5 jours.

Éloignement des bureaux de district

- Certains bureaux de district seront à plus d'un jour de route des bureaux de traitement. Le début des opérations de suivi peut être retardé à cause de longs délais d'exécution et de livraison.

La section qui suit traite les considérations stratégiques qui nous restent à analyser et à régler avant 1986 lorsque toutes les décisions concernant le traitement des données du recensement de 1990 seront prises.

Considérations stratégiques

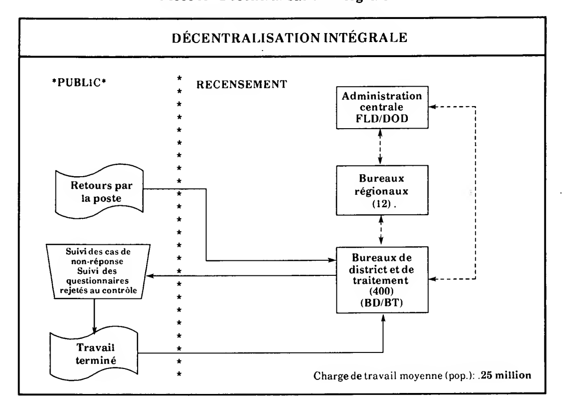
Les options de traitement des données du recensement de 1990 présentées ci-dessus sont aussi mises à l'épreuve au cours des essais officiels sur le terrain de 1985 et de 1986 visant à simuler les conditions de collecte et de traitement des données du recensement de 1990. Les sommes investies par le Census Bureau pour en savoir davantage sur les centres de traitement éloignés (1986 Census of Central Los Angeles County) et sur les opérations combinées de collecte et de traitement (1986 Census of East Central Mississippi) devraient fournir des renseignements utiles à la prise de décisions.

Les essais devraient nous aider de deux façons. Premièrement, ils nous permettront de vérifier nos modèles d'exploitation et d'adapter, selon les résultats obtenus, les hypothèses courantes concernant les taux de production, les coûts de maintenance et les difficultés de formation, et de nombreuses autres données. Deuxièmement, et surtout, ils nous permettront d'évaluer les impératifs de contrôle de gestion par rapport à la décentralisation. Cet aspect de l'exploitation de plusieurs centres de traitement est peut-être le plus inquiétant. Il soulève les questions du recrutement de 20 à 50 gestionnaires supérieurs, des locaux convenables, de l'enlèvement de matériel, de l'établissement de contrôles financiers et de la supervision des activités quotidiennes d'un réseau national de centres de traitement.

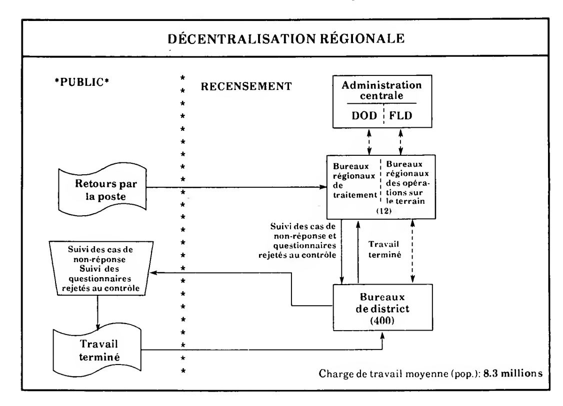
Comme le recensement décennal est un projet transitoire, il n'existe pas au Census Bureau d'infrastructure établie pour un vaste réseau de centres de traitement. En conséquence, les divisions du recensement décennal doivent compter sur les spécialistes du Census Bureau en traitement des données, en assistance technico-commerciale, en administration, en personnel et en approvisionnement. De plus, les pressions exercées sur ces groupes sont beaucoup plus grandes s'ils doivent assurer le soutien dans un contexte décentralisé plutôt que dans un cadre similaire à celui du recensement de 1980.

Donc, même si de nombreux facteurs qui militent pour ou contre la décentralisation feront l'objet d'analyses et d'essais rigoureux, le facteur déterminant de son succès ou de son échec sera plus difficile à évaluer. C'est pourquoi le Census Bureau doit soigneusement mettre au point une méthode de gestion de la saisie décentralisée des données qui réduise les risques tout en favorisant la réalisation des autres grands objectifs du recensement de 1990.

Pièce A - Décentralisation intégrale



Pièce B - Décentralisation régionale



UTILISATION DU CODAGE AUTOMATISÉ PAR STATISTIQUE SUÈDE

LARS LYBERG

SERVICE DE RECHERCHE STATISTIQUE
STATISTIQUE SUÈDE

Introduction

1. Activité de codage et aspects du problème de contrôle

Les opérations de traitement informatique qui se rattachent à une enquête comprennent, entre autres, la mise en formé, le codage, l'introduction en mémoire et la production de tableaux. Supposons une collection d'objets (ou d'"éléments") de nature quelconque et un ensemble de catégories qui s'excluent mutuellement. Chaque élément n'appartient qu'à une seule catégorie. Le codage consiste à classer les éléments dans les diverses catégories.

En pratique, le codage repose sur l'accès à l'information verbale concernant les éléments de la population ou de l'échantillon à l'étude. Cette information est habituellement consignée sur des questionnaires au cours de la collecte des données, par les répondants eux-mêmes ou par les interviewers ou les recenseurs. Contrairement à d'autres types d'information (données numériques sur les dépenses du ménage, par exemple), l'information verbale ne peut être consignée immédiatement sous forme de tableaux statistiques. Elle doit d'abord être codée par catégories et chaque catégorie doit être étiquetée au moyen, par exemple, d'un numéro. Ces numéros sont appelés numéros de code et la clé de ces numéros de code est appelée code. (Les données numériques peuvent évidemment être elles aussi codées; ainsi, dans un recensement du commerce, on peut classer les objets recensés dans des catégories définies en fonction du chiffre d'affaires, par exemple.)

Il est vrai que le terme "codage" est ambigu. On a tenté de le remplacer par le terme "classification". Ce terme est peut-être préférable à "codage", mais il présente des inconvénients. Dans la présente communication, nous utilisons le terme "codage" puisque c'est celui qu'on trouve le plus souvent dans les publications sur le sujet. Dans ce domaine, il y a d'autres termes qui sont tout aussi ambigus. Ainsi, ce que nous appelons "numéro de code" dans cette communication est souvent désigné par "code" dans les publications sur le sujet et ce que nous appelons ici "code" est souvent désigné par "liste de codes", "norme de codage", "lexique" ou "nomenclature". On ne s'entend pas non plus sur ce qu'il faut coder. Dans la définition ci-dessus, il est supposé qu'un élément donné appartient à une catégorie quelconque. Dans les documents sur le sujet, on définit souvent le codage comme une opération par laquelle les descriptions verbales des réponses sont codées, plutôt que les éléments eux-mêmes. Cette façon répandue de décrire la situation est facile à comprendre puisque, dans un grand nombre d'enquêtes, chaque élément est codé à l'égard de plus d'une variable.

L'activité de codage comporte les trois aspects suivants:

(1) Chaque élément d'une population, par exemple, doit être codé à l'égard d'une variable précise, à partir des descriptions verbales.

(2) Il existe un code pour cette variable, c'est-à-dire un ensemble de numéros de code désignant chacun une catégorie précise de la variable à l'étude.

(3) Une fonction de codage relie (1) et (2), c'est-à-dire un ensemble d'instructions de codage permet de relier les descriptions verbales aux numéros de code.

Le codage est une activité essentielle des études statistiques comme les recensements de la population, les recensements du commerce ou les enquêtes sur la population active. Quant aux variables, il peut s'agir de la profession, de l'industrie, de l'instruction et de la situation.

Le codage pose des problèmes de nature multiple. Comme la plupart des autres activités se rattachant aux enquêtes, le codage peut comporter des erreurs, car il arrive que la fonction de codage ne soit pas exécutée selon les règles et que la fonction de codage elle-même ou le code ne soient pas appropriés. En fait, dans certaines études statistiques, le codage constitue, après la collecte des

données, la principale source d'erreurs. Dans le cas de certaines variables, des taux de fréquence d'erreur de 10% ne sont pas rares. Le fait que le codage soit difficile à contrôler pose aussi un problème. Pour que le codage soit exact, le codeur doit faire preuve de beaucoup de jugement, car le choix du bon numéro de code peut être très difficile. Même parmi des codeurs expérimentés, le codage peut varier énormément. Il est donc difficile de concevoir des méthodes efficaces de contrôle de codage. Le troisième problème réside dans le fait qu'un grand nombre d'opérations de codage sont difficiles à gérer. Le codage a tendance à devenir une activité fastidieuse et coûteuse. Ainsi, lors du recensement de la population suédoise de 1970, le codage a nécessité plus de 300 années-personnes. Dans un grand nombre de pays, les codeurs affectés aux enquêtes de grande envergure doivent souvent être engagés de façon temporaire, avec les effets que l'on sait sur le niveau de la qualité. Il y a même lieu de croire qu'il sera peut-être difficile à l'avenir d'engager des codeurs temporaires pour ce genre de travail relativement monotone. Par conséquent, toute idée nouvelle permettant d'améliorer l'efficacité de l'activité de codage serait sûrement bien accueillie.

On trouve dans Lyberg (1981) un aperçu des problèmes relatifs au contrôle du codage.

2. Erreurs de codage en Suède

2.1 Recensement de la population suédoise de 1965

En 1967, on a procédé à une étude d'évaluation des erreurs de codage contenues dans le recensement de la population suédoise de 1965 (voir Lyberg (s.d.) et Dalenius et Lyberg (s.d.)). Sur un ensemble de questionnaires de recensement vérifiés représentant environ 70% de la population de 1965, on a prélevé un échantillon à deux degrés. La population a été divisée en quatre strates, et quatre sous-échantillons ont été obtenus. L'étude d'évaluation se limitait aux variables suivantes:

(1) Lien avec le chef de ménage
(2) Genre d'emploi
(3) Situation
(4) Industrie

Les codes utilisés pour les variables (1) à (3) ne comprenaient qu'un seul chiffre chacun, alors que le code utilisé pour l'"industrie" comprenait trois chiffres.

Comme nous travaillions avec quatre variables et quatre sous-échantillons, nous avons obtenu 16 estimations de taux d'erreur différentes, qui figurent au tableau 1.

Les sous-échantillons 1 et 3 se composaient de questionnaires de recensement intégralement vérifiés, alors que les sous-échantillons 2 et 4 se composaient de questionnaires de recensement vérifiés par prélèvement. La majeure partie de la vérification intégrale a été faite par des codeurs encore inexpérimentés. C'est ce qui explique les écarts de taux d'erreur entre les échantillons 1 et 3 et les échantillons 2 et 4.

2.2 Recensement de la population suédoise de 1970

Lors du recensement de la population suédoise de 1970, le nombre de variables à coder était supérieur à celui de 1965. Pour l'évaluation, on a tiré l'échantillon de l'ensemble des questionnaires de recensement. Un groupe de codeurs expérimentés a été chargé d'élaborer un ensemble de numéros de code d'évaluation "vrais" pour chaque questionnaire compris dans cet échantillon. Puis, on a comparé ces numéros de code aux numéros de code réels après la vérification pour obtenir des estimations des taux d'erreur relatifs aux

TABLEAU 1. Estimations des taux d'erreur (%) de codage réel lors du recensement de la population suédoise de 1965

Sous-échantillon	Variable			
	(1)	(2)	(3)	(4)
1	1.6	2.7	1.0	14.5
2	1.4	1.6	.6	8.2
3	1.5	3.0	1.3	14.5
4	.7	1.3	1.2	8.7

TABLEAU 2. Estimations des taux d'erreur relatifs au codage de l'activité économique lors du recensement de la population suédoise de 1970		
Variable	Code	Pourcentage du taux d'erreur (population globale)
(1)	1 chiffre	4.3
(2)	1 chiffre	4.7
(3)	3 chiffres	13.5
(4)	1 chiffre	3.7
(5)	4 chiffres	9.9
(6)	1 chiffre	8.9
(7)	1 chiffre	11.5
(8)	1 chiffre	4.4

différentes variables de l'activité économique. Les variables utilisées étaient les suivantes:

(1) Lien avec le chef de ménage
(2) Type d'activité
(3) Profession
(4) Situation
(5) Industrie
(6) Lieu de travail
(7) Moyen de transport utilisé pour se rendre au travail
(8) Nombre d'heures travaillées

Les estimations des taux d'erreur pour ces variables figurent au tableau 2.

Les taux d'erreur relatifs aux variables (1), (6) et (7) sont probablement exagérés puisque les numéros de code ont été traités par une machine de reconnaissance optique des caractères et qu'il y a lieu de croire que les erreurs techniques, à ce stade, n'ont qu'un effet négligeable sur les taux d'erreur relatifs à ces variables.

Le tableau montre qu'il est difficile de coder des variables comportant plusieurs chiffres. Cependant, il arrive assez souvent que même les variables d'un seul chiffre, qui sont considérées, à priori, comme faciles à coder, soient mal codées. Il se peut que le contexte de codage soit trop complexe pour un seul codeur, en ce sens que chaque codeur doit gérer un trop grand nombre de variables.

2.3 Recensement de la population suédoise de 1975

Lors du recensement de 1975, le nombre de variables était moins élevé qu'en 1970. Les études d'évaluation démontrent que les taux d'erreur étaient eux aussi moins élevés lors de ce recensement qu'en 1970. Les variables étudiées étaient les suivantes:

(1) Lien avec le chef de ménage
(2) Type d'activité
(3) Profession
(4) Situation
(5) Industrie
(6) Genre d'emploi
(7) Moyen de transport utilisé pour se rendre au travail

Toutes ces variables, à l'exception des variables (3) et (5), ne comprenaient qu'un seul chiffre. Le tableau 3 indique les taux d'erreur relatifs à ces variables.

TABLEAU 3. Estimations des taux d'erreur relatifs au codage de l'activité économique lors du recensement de la population suédoise de 1975		
Variable	Code	Pourcentage du taux d'erreur (population globale)
(1)	1 chiffre	.6
(2)	1 chiffre	.6
(3)	3 chiffres	7.8
(4)	1 chiffre	.5
(5)	4 chiffres	3.5
(6)	1 chiffre	1.0
(7)	1 chiffre	.5

Les résultats que contient le tableau 3 diffèrent sensiblement de ceux de l'étude d'évaluation de 1970. Les taux d'erreur ont chuté pour chacune des variables et il est très encourageant de voir que les variables d'un seul chiffre sont maintenant codées plus facilement. À 8%, le taux d'erreur relatif au codage de la profession est encore très élevé, mais comme il était de 13.5% en 1970, on peut dire que c'est un bon résultat. L'estimation relative à l'industrie est encore meilleure.

2.4 Autres études de Statistique Suède sur les taux d'erreur

La plupart des études de Statistique Suède sur le codage ont été menées dans le cadre de recensements, ce qui est plutôt normal puisque le codage est une activité essentielle de tout recensement. Au cours de la dernière décennie, on s'est intéressé de plus en plus aux erreurs de codage, ce qui a donné lieu à des études d'évaluation en rapport avec d'autres enquêtes également. Voici les taux d'erreur de codage calculés lors de ces études d'évaluation.

Dans Olofsson (1976), le taux d'erreur relatif à l'enquête sur la population active de 1974 est de 5.7%. Pour le codage de la profession, à l'intérieur de la même enquête, le taux d'erreur est de 6.2%. Dans Harvig (1973b), le taux d'erreur du codage de la profession, dans le cas des diplômés d'université, est évalué à 11%. Dans Harvig (1973a), le taux des erreurs de codage des causes de décès est évalué à 3.2%. Dans Lyberg et autres (1973), le taux des erreurs de codage du niveau d'instruction des enseignants est évalué à 8%. Dans ce cas, l'intervalle de confiance de 95% variait de 5.9% à 10.3%.

On trouve dans Lyberg (1983) un compte rendu détaillé des études relatives aux taux des erreurs de codage de l'industrie et de la profession.

3. Nécessité du contrôle

Dans la plupart des séries statistiques, il faut absolument que le contrôle du codage fasse partie du programme global de production des statistiques. Toutefois, pour planifier l'avenir, il ne suffit pas de connaître simplement le taux d'erreur. Il faut connaître la structure des erreurs, la fiabilité du processus de codage, les différents types d'erreurs, la gravité des erreurs et les répercussions des erreurs afin de prendre les mesures qui s'imposent à l'égard du code ou du codeur.

Il existe plusieurs moyens de contrôle:

Premièrement, il est possible, au moyen, par exemple, du modèle d'enquête du U.S. Bureau of the Census, d'analyser l'erreur survenue au cours d'une activité de codage particulière. Ce modèle aide aussi à bien équilibrer les différents contrôles exercés sur l'ensemble des activités d'enquête. (Voir Bailar et Dalenius (1969).)

Deuxièmement, le codage manuel se prête particulièrement bien à l'application des moyens statistiques de contrôle de la qualité élaborés, à l'origine, pour les applications industrielles. (Voir Minton (1969).)

Troisièmement, il existe certains moyens de contrôle conçus spécialement pour le codage, notamment la vérification connexe et la vérification indépendante. (Voir Lyberg (1981).)

Quatrièmement, l'évaluation des résultats de codage sert de base à la répartition des efforts de contrôle de la qualité. Nous avons déjà fourni des exemples de résultats de diverses études d'évaluation. Les résultats de ces études donnent une idée de la portée et de l'importance qu'il faut donner au programme de contrôle de qualité nécessaire.

Cinquièmement, on ne peut faire autrement que de remettre en question le principe même du codage manuel et d'examiner la possibilité d'élaborer une nouvelle approche au moyen de l'informatique. Cette idée n'est évidemment pas nouvelle en principe: ainsi, au U.S. Bureau of the Census, le codage géographique s'effectue au moyen d'ordinateurs depuis 1963. Ce qui est nouveau, c'est que cet organisme envisage d'utiliser beaucoup l'ordinateur pour le codage de variables aussi complexes que la profession et l'industrie. On peut considérer cette éventualité comme le prolongement naturel de l'application de l'informatique aux activités de contrôle.

Nous allons maintenant décrire les travaux suédois dans ce domaine précis d'automatisation.

4. Description sommaire du codage automatisé

Le codage automatisé comprend les quatre tâches suivantes:

(i) établissement d'un dictionnaire stocké en mémoire;
(ii) introduction des descriptions d'élément dans l'ordinateur;
(iii) appariement et codage;
(iv) évaluation.

4.1 Établissement d'un dictionnaire stocké en mémoire

Dans le codage automatisé, c'est le dictionnaire stocké en mémoire qui remplace la nomenclature et les instructions de codage utilisées pour le codage manuel. L'élaboration d'un tel dictionnaire constitue évidemment une tâche très importante. Le travail d'élaboration peut s'effectuer à la main, mais, comme il faut traiter des variables complexes comportant plusieurs chiffres, le recours à l'informatique semble préférable. Le dictionnaire ainsi constitué devrait renfermer un certain nombre de descriptions verbales et leurs numéros de code. Les descriptions peuvent consister en un échantillon des questionnaires à coder ou en un échantillon tiré d'une enquête précédente du même genre. La taille de l'échantillon utilisé pour constituer le dictionnaire pose, bien sûr, un problème important. Que le dictionnaire soit établi à la main ou au moyen d'un ordinateur, les numéros de code qu'il renferme devraient, en principe, être les codes attribués par les codeurs les plus compétents et être contrôlés au moyen de méthodes efficaces de vérification indépendante.

4.2 Introduction des descriptions d'élément dans l'ordinateur

Pour introduire les descriptions verbales dans l'ordinateur, on peut choisir l'introduction selon une structure plus ou moins imposée. Cette méthode présente cependant de sérieux inconvénients: elle nécessite beaucoup d'"espace", les erreurs résultant de l'introduction à grande échelle de l'information alphabétique sont relativement inconnues et, de plus, ce genre d'introduction en mémoire est généralement assez coûteux.

Il serait préférable de soumettre directement l'information verbale à la reconnaissance optique des caractères, mais, malheureusement, la reconnaissance des lettres manuscrites n'est pas encore assez avancée.

Il semble que l'introduction en mémoire des descriptions verbales soit actuellement le principal problème pratique auquel se heurte l'élaboration des systèmes de codage automatisés.

4.3 Appariement et codage

À l'heure actuelle, chaque description d'élément introduite dans l'ordinateur est comparée à la liste des descriptions d'activité professionnelle contenues dans le dictionnaire. Si une description d'élément concorde avec une description d'activité professionnelle (c'est-à-dire s'il y a "appariement"), elle reçoit le numéro de code correspondant; sinon, elle doit être codée manuellement.

Dans un système de codage automatisé, nous n'obtiendrons des concordances exactes que pour une fraction de tous les éléments. La première étape de l'élaboration d'un tel système doit donc consister à établir des critères relatifs au degré de similarité qui doit exister entre les mots introduits et les mots du dictionnaire pour qu'il y ait concordance.

4.4 Évaluation

Des études d'évaluation permanentes doivent être intégrées dans le système. De telles études visent les objectifs suivants:

(i) contrôle de la qualité du codage automatisé;
(ii) amélioration du dictionnaire; et
(iii) contrôle des coûts.

C'est le processus d'évaluation qui permet de déterminer si le codage automatisé est économique ou non. Les cas transmis par le service de codage manuel sont-ils plus difficiles à coder que ceux qui sont traités par l'ordinateur? Faut-il améliorer le dictionnaire? L'évaluation permet de résoudre toutes ces questions et bien d'autres.

5. Le dictionnaire

Il existe deux catégories générales d'algorithmes pour le codage automatisé: les

algorithmes de pondération et les algorithmes de dictionnaire. Les premiers attribuent des coefficients de pondération à chaque combinaison mot-code en utilisant l'information contenue dans un fichier de base; lorsqu'il doit coder un nouvel enregistrement, le programme choisit le numéro de code ayant le coefficient de pondération le plus élevé pour le mot en question. Les algorithmes de dictionnaire cherchent dans un dictionnaire les mots ou les chaînes de mots qui comportent des numéros de code précis. Lorsqu'il doit coder un nouvel enregistrement, le programme détermine si le mot ou la chaîne de mots correspond à un mot quelconque du dictionnaire. Si aucune concordance n'est repérée, l'enregistrement est rejeté et transmis au service de codage manuel.

Le U.S. Bureau of the Census a élaboré et étudié un certain nombre d'algorithmes au cours des dernières décennies. Pour certaines applications simples, comme le codage géographique, le codage automatisé s'est révélé très efficace. Les efforts plus récents portent surtout sur le codage plus complexe de la profession et de l'industrie. Quatre algorithmes sont décrits dans Lakatos (1977a, b). Deux d'entre eux, les algorithmes O'Reagan et Corbett, sont des algorithmes de dictionnaire. Les deux autres, l'IMP et l'INT, utilisent la méthode de pondération. L'algorithme INT, qui a été conçu par Rodger Knaus, est décrit en détail dans Knaus (s.d., 1978a, b, 1979, 1983). On trouve dans Appel et Hellerman (1983) et dans Appel et Scopp (1985) une description des travaux d'élaboration courants du U.S. Bureau of the Census.

À Statistique Suède, nous n'avons travaillé qu'avec des algorithmes de dictionnaire de sorte que nous n'avons rien à ajouter au sujet des autres algorithmes.

Par conséquent, le dictionnaire stocké en mémoire est l'équivalent du dictionnaire et des instructions de codage utilisés pour le codage manuel. L'établissement d'un tel dictionnaire suppose l'exécution des tâches suivantes:

(i) choix des données de base;
(ii) constitution d'un fichier de base par voie d'échantillonnage, à partir des données de base;
(iii) codage du fichier de base par des experts;
(iv) établissement des critères d'incorporation au dictionnaire;
(v) création d'un dictionnaire préliminaire;
(vi) vérification et parachèvement du dictionnaire préliminaire.

5.1 Choix des données de base

Les données de base les plus appropriées sont évidemment celles que renferment les déclarations remplies lors de l'enquête à l'étude, mais nous pouvons rarement les utiliser, car le temps nous manque. Ainsi, les données de base sont souvent:

(i) les données d'une ancienne enquête de nature analogue; ou
(ii) les données d'une enquête-témoin; ou
(iii) les données d'un autre genre d'enquête comportant la même variable.

Il faut cependant souligner que des données de base appropriées, du genre mentionné ci-dessus, peuvent être utilisées efficacement pour la révision d'un dictionnaire exploité depuis déjà un certain temps.

Il est indispensable que les données de base soient à jour. Les changements structuraux se produisent à l'intérieur d'une population; ainsi, la création ou la suppression d'appellations d'industrie et de profession n'est pas rare. De plus, il se peut que la façon de répondre des répondants change avec le temps, comme le démontre l'exemple suivant. Lors du recensement de la population suédoise de 1965, les nettoyeurs utilisaient le mot "nettoyeur" pour décrire leur activité professionnelle. Lors du recensement de 1970, certains nettoyeurs ont utilisé l'expression "préposé à l'entretien des locaux". Mais comme cette expression n'existait pas en 1965, elle ne faisait pas partie des données de base. Le dictionnaire basé sur les données du recensement de 1965 n'était donc pas en mesure de coder les répondants qui, au recensement de 1970, ont décrit leur activité professionnelle par l'expression "préposé à l'entretien des locaux".

Les données de base de la catégorie (iii) ne devraient être utilisées que dans des circonstances exceptionnelles puisque les réponses relatives à une certaine variable peuvent différer sensiblement d'une enquête à une autre en raison de modes différents de collecte des données.

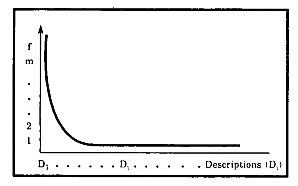
5.2 Constitution d'un fichier de base par voie d'échantillonnage, à partir des données de base

Il faut, à partir des données de base, prélever un échantillon d'enregistrements pour constituer un dictionnaire. L'échantillon d'enregistrements peut être constitué des façons suivantes:

- échantillon simple pris au hasard,
- échantillon au hasard contrôlé, ou
- échantillon subjectif.

Selon la première méthode, les descriptions dont le taux de fréquence est bas ont peu de chance de figurer dans le fichier, ce qui n'est généralement pas une conséquence négative. Nous avons donc eu recours à l'échantillon simple pris au hasard pour presque tous nos travaux et toutes nos applications. La taille de l'échantillon pose un problème, quelle que soit la méthode utilisée, puisqu'il faudrait que chaque description soit codée par des "experts".

Pour certaines de ses expériences de codage automatisé, le U.S. Bureau of the Census a choisi un très large échantillon initial d'enregistrements: des échantillons d'environ 100,000 enregistrements ont été utilisés. Pour ses expériences, Statistique Suède a utilisé un fichier de base ne renfermant pas plus de 14,000 enregistrements. Malgré cela, les études d'évaluation ont révélé des résultats comparables à ceux du Bureau of the Census, que l'on peut expliquer par le fait que quelques numéros de code et quelques descriptions de dictionnaire suffisent, pour un grand nombre de variables, à coder une grande partie des enregistrements. De plus, la langue suédoise est moins complexe (du moins dans ce contexte) que l'anglais. Voici un diagramme des fréquences pour types de descriptions uniques:



Le diagramme type possède une queue très irrégulière si les descriptions sont classées selon leur fréquence d'occurrence. En fait, dans certaines applications, un grand nombre de descriptions uniques ne surviennent qu'une ou deux fois. D'après O'Reagan (1972), une étude plus approfondie a révélé que, pour une variable, 7% des numéros de code pouvaient traiter 50% des enregistrements. Ainsi, il est généralement possible de constituer un dictionnaire acceptable au moyen d'un échantillon initial plutôt petit. D'après notre expérience, le fait d'élargir sensiblement le fichier de base (une fois qu'il satisfait aux critères d'acceptation) n'améliore pas beaucoup le niveau de codage. Il semble plus efficace de concentrer ses efforts sur les catégories les plus fréquemment utilisées et de recourir, pour le reste, au codage manuel.

5.3 Codage du fichier de base par des experts

Pour constituer un bon dictionnaire, il faut que le fichier de base soit codé avec une grande exactitude. Pour ce faire, il faut utiliser les meilleurs codeurs possibles. Comme même le codage effectué par des "experts" est susceptible d'erreurs, le codage hautement exact du fichier de base doit s'accompagner d'un processus de contrôle.

5.4 Établissement des critères d'incorporation au dictionnaire

Les descriptions verbales que renferme un fichier de base codé par des experts peuvent appartenir à plusieurs catégories:

a) descriptions très fréquentes qui sont toutes associées à un numéro de code précis;
b) descriptions peu fréquentes qui sont toutes associées à un numéro de code précis;
c) descriptions dont la fréquence est basse ou élevée et qui sont rattachées à des numéros de code différents.

En principe, il faudrait inclure dans le dictionnaire toutes les descriptions se rattachant à un numéro de code précis, mais, dans la pratique, tout dépend de la taille du dictionnaire que l'on juge acceptable. La taille acceptable est liée, à son tour, au temps de recherche du programme d'appariement. Si le temps de

recherche est indépendant de la taille du dictionnaire et si la création d'un dictionnaire de grande taille ne nécessite pas un travail de gestion manuel trop important, il faut inclure toutes les descriptions se rattachant à des numéros de code précis. Sinon, il faut définir ce qu'on entend par "fréquence élevée". Cette décision dépend, entre autres, de la taille de l'échantillon et du nombre de catégories que renferme le code: ainsi, si l'échantillon de base est passablement restreint, il n'y aura pas de descriptions très fréquentes. À notre avis, il est préférable d'attribuer à la notion de "fréquence élevée" f une valeur basse, comme $f \geq 3$, puisqu'il est généralement plus facile de supprimer des descriptions du dictionnaire que d'en ajouter.

Les descriptions appartenant à la catégorie c) ne devraient pas faire partie du dictionnaire. Il existe cependant des exceptions à cette règle. Si, dans une grande majorité de cas, une description très fréquente est associée à un numéro de code précis, on peut envisager d'inclure cette description dans le dictionnaire. Si on inclut une telle description, on se retrouve évidemment avec une classification erronée, délibérément incorporée au dictionnaire. Même s'il est vrai que de tels taux d'erreur soient faibles, il est probablement préférable de modifier la nomenclature pour que le codage de cette description particulière soit d'abord exempt d'ambiguïté.

5.5 Création, vérification et parachèvement du dictionnaire préliminaire

La création d'un dictionnaire peut se faire manuellement ou par ordinateur. Pour plus d'efficacité, il est probablement préférable de combiner ces deux méthodes. Lors de nos premières expériences à Statistique Suède, nous avons utilisé des dictionnaires constitués à la main, mais nous pouvons maintenant utiliser un programme informatique de création de dictionnaire.

On peut dire que la création manuelle des dictionnaires se fait par approximations successives. À Statistique Suède, nous avons utilisé deux listes: la liste nº 1 est le fichier des descriptions codées par des experts et triées par numéros de code, et la liste nº 2, le même fichier trié par ordre alphabétique. La création du dictionnaire se fait à partir de ces listes. La liste nº 1 fournit des précisions sur la structure des descriptions verbales classées sous un numéro de code particulier. Nous choisissons une limite de fréquence f pour définir les descriptions "très fréquentes". Toutes les descriptions relevées f fois ou plus sont stockées dans la version préliminaire du dictionnaire principal qui, au moment du codage automatisé, est consulté en premier. Ce dictionnaire porte le nom de PLEX.

Pour augmenter le degré de codage, il faut inclure certaines variantes des descriptions très fréquentes déjà stockées. On peut, pour ce faire, reconnaître des chaînes de mots discriminantes. Dans la situation idéale, une telle chaîne représente de nombreuses variantes d'une même description. Ainsi, après avoir stocké les descriptions très fréquentes, on recherche les chaînes de mots discriminantes. Ces chaînes (ou plutôt ces parties de mot) sont stockées dans un dictionnaire auxiliaire. Ce dictionnaire auxiliaire, appelé SLEX, est consulté si le PLEX ne réussit pas à coder la description.

La liste nº 2 sert de liste de contrôle. Une description préliminaire stockée dans le PLEX a-t-elle reçue un autre numéro de code, en plus du numéro de code à l'étude? Il est fréquent qu'une description soit associée à différents numéros de code, selon le code, les instructions de codage et l'information connexe utilisés par les codeurs. La liste alphabétique aide à repérer ces descriptions. On peut ensuite les omettre du dictionnaire PLEX préliminaire, de même qu'omettre les chaînes de mots connexes du dictionnaire SLEX préliminaire. Toutefois, comme nous l'avons mentionné ci-dessus, si nous permettons délibérément un certain degré de codage erroné, certaines de ces descriptions ambiguës demeureront dans le dictionnaire. L'éventualité d'une telle classification erronée est cependant faible.

Il arrive que des descriptions très fréquentes soient omises parce qu'elles sont ambiguës. On peut alors envisager d'incorporer dans le PLEX les descriptions peu fréquentes mais non équivoques, ou encore, de transformer certaines descriptions ambiguës en descriptions non équivoques à l'aide de l'information connexe.

Les chaînes de mots du SLEX doivent être communes à plusieurs descriptions ou représenter des parties de descriptions spéciales dont la fréquence est élevée. Il faut être sûr que les mots du SLEX ne correspondent pas à des descriptions du PLEX rattachées à d'autres numéros de code. Le SLEX ne peut jamais être élargi, à cause des problèmes d'exactitude qu'il pose. La principale difficulté avec le SLEX, c'est que nous ne savons pas d'avance comment il se comporte lorsque de nouveaux enregistrements sont codés.

L'activité manuelle décrite ci-dessus (ou toute autre méthode manuelle analogue) peut, dans une large mesure, être exécutée par l'ordinateur. On trouve, dans O'Reagan (1972) (algorithme O'Reagan) et dans Corbett (1972) et Owens (1975) (algorithme de Corbett), deux méthodes élaborées par le U.S. Bureau of the Census.

Le système informatisé de création de dictionnaire qu'utilise Statistique Suède génère un dictionnaire à deux volets, le PLEX et le SLEX. Le PLEX renferme des descriptions non équivoques et est consulté en premier. Le SLEX renferme des chaînes de mots discriminantes qui sont communes à plusieurs descriptions introduites. Par conséquent, le SLEX n'est pas aussi exact que le PLEX et il n'est consulté que si le PLEX ne réussit pas à coder les descriptions. D'après notre expérience, il est plutôt facile de créer un PLEX manuellement, mais la création manuelle du SLEX est beaucoup plus difficile. Par conséquent, nous avons élaboré un programme de création automatisée du SLEX. (La création d'un PLEX informatisé ne devient alors qu'un cas spécial simple.)

Nous avons mis à l'essai quelques versions différentes de ce programme. La version actuelle (un progiciel appelé AUTOCOD) est décrite dans Bäcklund (1978). Tous les programmes sont écrits en PL1. Le progiciel AUTOCOD renferme des sous-programmes pouvant effectuer les tâches suivantes:

- création de dictionnaires stockés en mémoire d'ordinateur (PCLEXK)
- codage des descriptions (PCAUTOK)
- mise à jour des dictionnaires (PCLEXUP)
- évaluation des dictionnaires (PCLEXT)

Le PCLEXK génère un PLEX et un SLEX. Cette procédure comporte trois étapes. Le programme LEXLADD génère de l'espace pour un SLEX éventuel. Le LEXKONS crée le PLEX et le SLEX. Pour chaque description PLEX, le système vérifie une abréviation de six caractères, par exemple, en commençant par le premier caractère, en vue de l'incorporer au SLEX. Si cette abréviation correspond à une autre description contenue dans le PLEX, elle est rejetée et une nouvelle abréviation est créée à partir du deuxième caractère de la description PLEX. Cette procédure est répétée au plus six fois; si aucune abréviation valide n'est obtenue, le programme passe à la description PLEX suivante. Enfin, le LEXLIST liste le contenu des dictionnaires au moyen d'EASYLIST. Les paramètres qui peuvent prendre des valeurs variables sont les suivants:

- utilisation possible d'une liste de préfixes qui, lorsqu'on crée un dictionnaire de biens, par exemple, supprime les chaînes de mots comme "livres", "rouleau" et "paires";
- fréquence minimale f_0 (c'est-à-dire le critère d'incorporation au dictionnaire);
- degré d'ambiguïté toléré;
- longueur minimale des mots contenus dans le SLEX.

Le PCAUTOK code les nouvelles descriptions au moyen du PLEX et du SLEX. On utilise le PCLEXUP pour supprimer les descriptions d'un dictionnaire ou pour en ajouter. Le PCLEXT sert à évaluer un dictionnaire lorsque nous avons accès à des données auxquelles des numéros de code ont été attribués manuellement.

Le PLEX et le SLEX peuvent être mis à jour simultanément ou séparément.

6. Utilisation de l'information connexe

Lorsque nous effectuons le codage manuellement, il nous arrive souvent d'utiliser non seulement les descriptions verbales pour la variable à coder, mais aussi d'autres catégories d'information connexe, c'est-à-dire des descriptions se rattachant à des variables connexes. Ainsi, l'information sur l'instruction ou l'industrie sert parfois d'information connexe pour le codage de l'activité professionnelle.

Il est évident que l'information connexe peut aussi être utilisée pour le codage automatisé. Il faut non seulement que cette information connexe accompagne les descriptions à coder, mais aussi qu'elle figure dans le dictionnaire. Le stockage de l'information connexe dans le dictionnaire et l'élaboration des programmes informatiques d'appariement et de codage ne posent aucun problème sérieux. En outre, l'information connexe peut être utilisée efficacement si le codage est exécuté en deux étapes; les résultats obtenus au cours de la première étape de codage peuvent servir d'information connexe à la seconde étape. Si les variables de la première étape sont codées manuellement, les numéros de code peuvent être introduits avec les descriptions verbales des variables à coder au cours de la seconde étape. Comme l'introduction des descriptions verbales est une tâche fastidieuse, la sortie des résultats de la première étape peut se faire plus rapidement. Les économies de temps ainsi réalisées dans le cas d'une enquête de l'envergure d'un recensement démographique peuvent être considérables; surtout si les variables de la seconde étape sont difficiles à coder. C'est ce qui s'est produit pour le codage de la profession lors du recensement de la population de 1980.

7. Évaluation et contrôle

L'étape finale et nécessaire au codage automatisé est celle de l'évaluation et du contrôle. Elle vise surtout à maintenir le niveau d'exactitude préalablement spécifié.

Le degré de codage, p, et la proportion correctement codée, q, représentent les principales caractéristiques vérifiées au moment du contrôle et de l'évaluation. Si N représente le nombre d'éléments introduits dans l'ordinateur et n, le nombre réellement codé, alors $p = n/N$. Si m des n éléments à coder ont été correctement codés, alors $q = m/n$. L'évaluation d'une procédure de codage automatisé doit porter à la fois sur p et sur q. Il est évidemment plus important d'obtenir une valeur q élevée qu'une valeur p élevée. Pour la comparaison de résultats différents, le produit pq peut se révéler utile. Malheureusement, il faut utiliser cette mesure avec beaucoup de précautions. Ainsi, la combinaison $p = .5$, $q = .9$ est de beaucoup supérieure à la combinaison $p = .9$, $q = .5$. On doit viser surtout la valeur q la plus élevée possible. Une fois cette valeur obtenue, on peut s'efforcer d'augmenter la valeur p. Cette proportion peut être augmentée jusqu'à ce que q commence à décroître. Il est même possible d'augmenter la valeur p aux dépens de la valeur q, pour autant que les avantages monétaires compensent largement la baisse du niveau de qualité.

Le coût du codage manuel de la proportion 1-p joue un rôle important dans le calcul des coûts du codage global, y compris les étapes automatisées et manuelles. Les descriptions que l'ordinateur ne réussit pas à coder peuvent être plus complexes que celles qu'il réussit à coder. De plus, les coûts fixes du codage manuel de la proportion 1-p sont relativement plus élevés que ceux du codage manuel de tous les éléments et, qui plus est, tous les numéros de code manuel doivent être introduits en mémoire. L'évaluation du codage automatisé doit tenir compte de ces coûts. Toutefois, des expériences récentes démontrent que, dans le cas d'un recensement, la proportion 1-p peut être codée en grande partie sans accès aux questionnaires, ce qui rend le processus plus rapide que le codage manuel classique.

Le second objectif du processus d'évaluation et de contrôle est de réunir les données qui pourront servir à modifier les dictionnaires et les programmes d'appariement. Des échantillons peuvent être prélevés de façon continue parmi les résultats produits pour être codés par des vérificateurs expérimentés selon une méthode appropriée (par voie de vérification indépendante, par exemple). Il est donc possible d'évaluer continuellement q. Si q ne satisfait pas aux normes de qualité, il faut analyser l'échantillon à l'étude. Quels sont les types de descriptions codés par erreur et quels sont ceux qui n'ont pas été codés du tout? L'échantillon est-il exceptionnel sous un aspect quelconque? Est-ce que des sections spéciales du code posent des problèmes à l'ordinateur? On peut, bien sûr, se poser toutes ces questions, et bien d'autres, même si la valeur-q réelle est conforme aux normes préalablement spécifiées. Mais, en règle générale, il ne faudrait apporter des rectificatifs que lorsque le processus est déréglé ou menace de l'être.

Expériences

Avec les années, Statistique Suède a beaucoup expérimenté le codage automatisé. À la longue, les résultats de nos expériences sont devenus si concluants que nous avons décidé d'adopter cette méthode pour nos enquêtes courantes. Dans la présente section, nous allons présenter certains aspects généraux de nos premières expériences en espérant qu'ils seront utiles à d'autres chercheurs.

8. Industrie

Les toutes premières expériences de codage automatisé à Statistique Suède portaient sur la variable "industrie". Ces expériences n'ont pas été très fructueuses. Les valeurs-q ne dépassaient pas .83 et, lors d'un essai particulier, la valeur-q n'était que de .69. Au cours de cette première phase d'expérimentation, nous pensions qu'un système de codage automatisé se devait d'être plutôt complexe, et même compliqué. Nous étions convaincus qu'un codage reposant uniquement sur l'appariement exact donnerait des résultats très décevants. À notre avis, la structure de réponse est si complexe que le degré de codage obtenu selon des règles d'appariement exact serait trop faible pour être rentable. Nous avons donc mis à l'essai des règles d'appariement spéciales. Nous avons tenté de mesurer l'"écart" entre les descriptions des répondants et les descriptions contenues dans le dictionnaire. Nous avons aussi essayé d'utiliser le coefficient de corrélation des rangs de Spearman pour mesurer la similarité entre ces deux sortes de descriptions. Des expériences plus récentes démontrent que, compte tenu du niveau d'exigences, ceci n'est peut-être pas nécessaire.

Après ces essais initiaux, nous n'avons pas travaillé beaucoup avec la variable "industrie". En fait, la majeure partie du travail par "tâtonnement", qui, à notre avis, est essentiel à l'élaboration des méthodes de codage automatisé, reste encore à faire en ce qui concerne cette variable. On pourrait même avancer que, la plupart du temps, les descriptions de l'industrie, sans consultation de l'information connexe, sont plus ou moins inutiles pour les codeurs manuels également. Par conséquent, le codage automatisé de cette variable ne peut se faire uniquement à partir des descriptions de l'industrie. Comme nous avions hâte de présenter des résultats à ceux qui subventionnent nos recherches, nous avons commencé à nous tourner dans une autre direction: vers la variable "profession".

9. Profession

La majeure partie de nos expériences portait sur la variable "profession". Le tableau 4 offre un résumé de nos principaux travaux.

Comme le montre ce tableau, les valeurs-q variaient de .85 à .95, résultat que nous jugions excellent. Pendant cette période d'essai, le programme de création automatisé de dictionnaires a été amélioré. Les règles complexes d'appariement pouvaient être utilisées sur demande, la taille du dictionnaire variait de 900 à 11,000 enregistrements et nous nous sommes aperçus qu'il fallait

TABLEAU 4. Expériences de codage automatisé de la profession

Expérience	Genre de dictionnaire	Enquête	Degré de codage (%)	Qualité ou taux de concordance (%)
1	Manuel	Recensement de 1965	62	95
2	Manuel	Recensement de 1970	66	92
3	Manuel	Recensement de 1970	74	84
4	Manuel	Recensement de 1970	80	90
5	Manuel	Enquête sur la population active de 1974	81	81
6	Automatisé (PLEX + SLEX)	Recensement de 1970	69	87
7	Automatisé (PLEX + SLEX)	Enquête sur la population active de 1976	84	85
8	Automatisé (SLEX)	Enquête sur la population active de 1976	69	93
9	Automatisé (PLEX)	Enquête sur la population active de 1976	69	92
10	Manuel et automatisé	Enquête sur la population active de 1976	74-76	93-94

utiliser le SLEX avec beaucoup de précautions. La qualité n'augmente jamais grâce au SLEX. Le degré de codage augmente, bien sûr, mais toujours aux dépens de la qualité. Ainsi, lors de l'expérience 7, le niveau de la qualité est passé de 84 à 92 grâce à l'utilisation du PLEX seulement, mais le degré de codage est passé de .84 à .69. Il est intéressant de noter que toute modification légère du PLEX n'a aucun effet sur le degré de codage et sur la qualité. Toutefois, il existe une liste alphabétique d'environ 11,000 descriptions de profession. Cette liste utilisée par les codeurs manuels **n'est pas** fondée sur la connaissance des structures de réponse empiriques. Nous voulions savoir ce qui arriverait si nous utilisions cette liste, qui était déjà sur bande. Nous l'avons utilisé lors des expériences 8 et 9 et nous avons obtenu des degrés de codage de 40.2% et 36.0% respectivement. Un document de ce genre ne peut donc pas se comparer à un PLEX établi en fonction de structures de réponse empiriques, même si le PLEX utilisé dans cette étude ne renfermait que 1,637 descriptions. Lorsque nous avons fusionné notre PLEX informatisé et cette liste alphabétique pour constituer un PLEX élargi, le degré de codage a augmenté jusqu'à environ 75%. Une telle combinaison peut donc être utile.

10. Biens

Les biens ou les achats représentent une des principales variables de toute enquête portant sur les dépenses du ménage. Le codage de cette variable est relativement simple en comparaison, par exemple, du codage de la profession. Le cas ordinaire n'est pas compliqué. Les fréquences d'erreur de codage observées lors d'enquêtes portant sur les dépenses du ménage confirment cette assertion. Les travaux portant sur cette variable, exécutés uniquement à l'aide du PLEX, présentent des valeurs-q d'environ .995 et des degrés de codage d'environ .68. Comme ce codage est simple mais coûteux et fastidieux, il se prête particulièrement bien à l'automatisation.

Applications

Statistique Suède a utilisé le codage automatisé pour un certain nombre d'enquêtes réelles. La toute première application a été le codage des biens lors de l'Enquête sur les dépenses du ménage de 1978. On a ensuite eu recours au codage automatisé pour effectuer le codage de la profession et de la catégorie socio-économique (SEI) lors du recensement de la population de 1980, ainsi que pour le codage des biens lors de l'Enquête sur les dépenses du ménage de 1985. On a aussi utilisé cette méthode pour coder les auteurs et les titres de livres dans le cadre d'une enquête continue sur les prêts de livres, pour coder la profession et la catégorie socio-économique dans le cadre de l'enquête (continue) sur les conditions de vie et pour coder l'activité professionnelle dans les enquêtes sur les étudiants. D'autres applications sont en cours de préparation.

11. Codage des biens lors des Enquêtes sur les dépenses du ménage de 1978 et de 1985

11.1 Introduction

Pour l'Enquête sur les dépenses du ménage de 1978, 5,900 ménages étaient censés tenir un registre détaillé de leurs achats de biens au cours d'une période de deux semaines. Le reste de l'échantillon, soit environ 7,900 ménages, devait tenir un registre simplifié des biens achetés au cours d'une période de quatre semaines.

Pour l'Enquête sur les dépenses du ménage de 1985, 6,000 ménages étaient censés tenir un registre des biens achetés au cours d'une période de quatre semaines, la complexité du registre à tenir se situant entre celle du registre détaillé et celle du registre simplifié de 1978. Pour cette enquête, qui n'est pas encore parachevée, le même numéro de code a été attribué à tous les produits alimentaires.

Pour ce genre d'enquête, les répondants peuvent faire parvenir leur registre pendant toute la durée de l'année sur laquelle porte l'enquête. Les données peuvent ainsi être traitées par cycles, ce qui peut être avantageux dans un système de codage automatisé. En 1978, comme nous n'étions pas du tout persuadés que notre système fonctionnerait dans un cadre d'exploitation, nous avons décidé, pour commencer, d'effectuer le codage au moyen de deux systèmes parallèles, l'un manuel et l'autre automatisé. Après deux mois de travail, nous avons procédé à l'évaluation des deux systèmes afin d'en choisir un qui serait utilisé au cours des 10 mois suivants, et c'est le codage automatisé qui est sorti gagnant de l'épreuve. Pour l'enquête de 1985, nous avons eu recours au codage automatisé dès le départ.

11.2 Le système automatisé

En 1978, la création du dictionnaire s'est faite par étapes. L'établissement du dictionnaire initial a nécessité des efforts considérables, puis des révisions ont continuellement été effectuées avant un grand nombre de cycles. Le dictionnaire initial se fondait sur le dictionnaire utilisé lors des expériences déjà mentionnées sous la rubrique "Expériences", ainsi que sur une liste de toutes les descriptions faisant partie des données expérimentales. Chaque description unique a été codée par des experts en enquêtes sur les dépenses du ménage. La création du dictionnaire a dû être effectuée en grande partie manuellement puisque la nature des descriptions avait changé au cours des neuf années qui s'étaient écoulées depuis l'Enquête sur les dépenses du ménage de 1969 dont nous avions tiré les données expérimentales. Seul un PLEX a été constitué, avec un taux de descriptions non ambiguës de 100%. Ce dictionnaire initial renfermait 1,459 descriptions. Lors du codage automatisé, les descriptions non codées ont été listées alphabétiquement sous une forme appropriée à la reconnaissance optique des caractères. Les numéros de code ont été ajoutés directement à cette liste. Certaines des descriptions non codées ont été incorporées au dictionnaire au moment de la mise à jour.

En 1985, on s'est servi du dictionnaire de 1978 (c'est-à-dire de la dernière version contenant 4,230 descriptions) comme point de départ. Cette partie **manuelle** du travail a donné un premier dictionnaire contenant 1,985 descriptions PLEX. À l'heure actuelle, l'Enquête sur les dépenses du ménage de 1985 se poursuit toujours.

11.3 Résultats

Au cours de l'enquête de 1978, 33 cycles ont été exécutés. Pendant cette période, 17 versions différentes du PLEX ont été utilisées; par conséquent, seules les données de quelques cycles ont été codées à l'aide de dictionnaires identiques. Le tableau 5 ci-dessous indique la taille du dictionnaire et le degré de codage pour chacun des cycles.

TABLEAU 5. Taille du dictionnaire et degré de codage pour les 33 cycles de l'Enquête sur les dépenses du ménage de 1978

Version du dictionnaire	Cycle	Nombre de descriptions contenues dans le dictionnaire	Degré de codage (%)
1	1	1,459	56
2	2	1,554	63
3	3	1,760	67
4	4	2,228	66
5	5, 6, 7	2,464	68, 68, 63
6	8	1,632	64
7	9	1,990	53
8	10, 11	2,451	69, 66
9	12	2,866	61
10	13	3,065	68
11	14	3,613	58
12	15, 16	3,752	72, 73
13	17, 18	3,832	39, 70
14	19, 20	4,011	65, 73
15	21, 22	4,229	51, 72
16	23, 24, 25, 26,	4,230	64, 67, 62, 67,
	27, 28, 29, 30		72, 65, 67, 50
17	31, 32, 33	4,230	65, 39, 67

Le degré de codage pour l'ensemble des cycles était de 65%. Comme nous pouvons le voir d'après le tableau, le degré de codage fléchit parfois sensiblement. Ceci s'explique par le fait que les registres détaillés sont plus faciles à coder automatiquement que les registres simplifiés et que la proportion de registres détaillés varie d'un cycle à un autre. En fait, certains cycles ne comportaient qu'un seul type de registre. Le degré de codage est évalué à 70% pour les registres détaillés et à 38% pour les registres simplifiés.

Le dictionnaire a subi des modifications avant la plupart des cycles de traitement, du moins pendant la majeure partie de la phase d'exploitation. Comme l'indique le tableau 5, le nombre d'additions dépasse généralement le nombre de suppressions. Ces modifications n'ont cependant pas influé beaucoup sur le degré de codage. Un examen plus attentif révèle qu'un grand nombre de mots du dictionnaire ont servi rarement, ou pas du tout, et qu'un nombre relativement faible de mots du dictionnaire ont servi à la plupart des descriptions introduites. Toutes les nouvelles descriptions non codées ont été réunies dans un fichier spécial. Dès qu'une nouvelle description apparaissait dans les réponses d'au moins trois ménages, elle était incorporée dans le dictionnaire pourvu qu'elle puisse être classée de façon non ambiguë dans une catégorie précise.

On utilise essentiellement la même procédure pour l'Enquête sur les dépenses du ménage de 1985. Le tableau 6 montre la taille du dictionnaire et le degré de codage pour les 17 cycles exécutés jusqu'ici.

Pour ces enquêtes sur les dépenses du ménage, il a été décidé d'utiliser uniquement le PLEX en raison de l'inefficacité du fichier SLEX. D'après nos calculs, le degré de codage baisse de 10 à 15 points lorsque le SLEX n'est pas utilisé. Cependant, pour la proportion de 65% qui est codée, la qualité de codage est élevée, le taux d'erreur étant inférieur à 1%. Des études d'évaluation spéciales ont démontré que la qualité du codage, pour le reste, était également excellente, le taux d'erreur s'élevant à 1% environ. On ne peut absolument pas comparer ce taux à celui qu'on obtiendrait avec le SLEX. En somme, le codage pour l'Enquête sur les dépenses du ménage de 1978 s'est fait sans problème. Les opérateurs ont trouvé qu'il était moins ennuyeux d'introduire, cette fois, des

TABLEAU 6. Taille du dictionnaire et degré de codage pour les 17 cycles exécutés jusqu'ici (1er septembre 1985) dans l'Enquête sur les dépenses du ménage de 1985

Version du dictionnaire	Cycle	Nombre de descriptions contenues dans le dictionnaire	Degré de codage (%)
1	1, 2	1,985	81, 78
2	3	2,029	82
3	4, 5	2,063	82, 81
4	6	2,126	82
5	7	2,156	83
6	8	2,176	82
7	9	2,207	81
8	10	2,228	83
9	11, 12, 13	2,272	83, 83, 82
10	14, 15, 16	2,370	83, 80, 81
11	17	2,446	83

déclarations verbales. On a aussi calculé que le codage automatisé coûtait entre 2 et 5% de moins que le codage effectué selon les méthodes manuelles classiques. Il présente aussi certains autres avantages: puisque toutes les descriptions sont introduites en mémoire, les données principales sont mieux étayées que si seul le numéro de code était introduit. Par conséquent, il est possible de fournir des descriptions plus détaillées des biens contenus dans les groupes pour lesquels des estimations sont fournies. En outre, comme le dictionnaire réussit à coder la plupart des descriptions simples, le reste du codage devant être effectué à la main devient plus intéressant pour le codeur. Le codage pour l'Enquête sur les dépenses du ménage de 1985 se poursuit lui aussi sans problème. Nous avons tiré parti de l'expérience acquise en 1978 et avons réduit les tâches administratives.

Comme l'indique le tableau 5, il ne semble pas utile d'apporter des modifications profondes au dictionnaire, passé un certain point. On obtient très tôt un degré de codage qu'on ne peut modifier sensiblement sans changer également la structure du dictionnaire. Dans le cas de l'enquête de 1978, nous avons obtenu un degré de codage de 67% dès la troisième version. En dépit d'efforts considérables et de modifications nombreuses passé ce point, nous n'avons pu obtenir un taux dépassant 73%.

Jusqu'ici, l'enquête de 1985 semble évoluer de la même façon. Le dictionnaire n'a pas encore atteint une taille aussi grande qu'en 1978, mais l'extension réelle dont il a été l'objet (de 1,985 à 2,446 descriptions) n'a pas influé sur le degré de codage, qui s'est stabilisé entre 81 et 83%.

Le codage automatisé est très économique. Le codage de dizaines de milliers de descriptions coûte moins de 50 dollars canadiens. Les frais informatiques de la mise à jour des dictionnaires sont à peu près les mêmes. Ce sont la préparation manuelle et l'administration qui coûtent cher. Comme nous l'avons mentionné, nous nous sommes acquittés de ces tâches de façon plus efficace dans l'enquête de 1985.

12. Codage automatisé de la profession et de la catégorie socio-économique dans le recensement de la population de 1980

12.1 Introduction

Au recensement de la population de 1980, le codage de la profession et de la classification socio-économique (SEI) est automatisé, ce qui signifie que les données d'identification personnelles et les descriptions de profession sont introduites et comparées à celles que renferme le dictionnaire stocké en mémoire. Le dictionnaire renferme les descriptions de profession accompagnées des numéros de code de profession et de SEI correspondants.

Le code de profession utilisé dans le recensement est structuré d'après la "Northern Standard for Classification of Occupations" (NYK) qui, à son tour, se fonde sur la "Classification internationale type des professions" (CITP). Ce code renferme environ 280 catégories distinctes de trois chiffres. Le code SEI est un code de deux chiffres qui comporte 14 catégories distinctes.

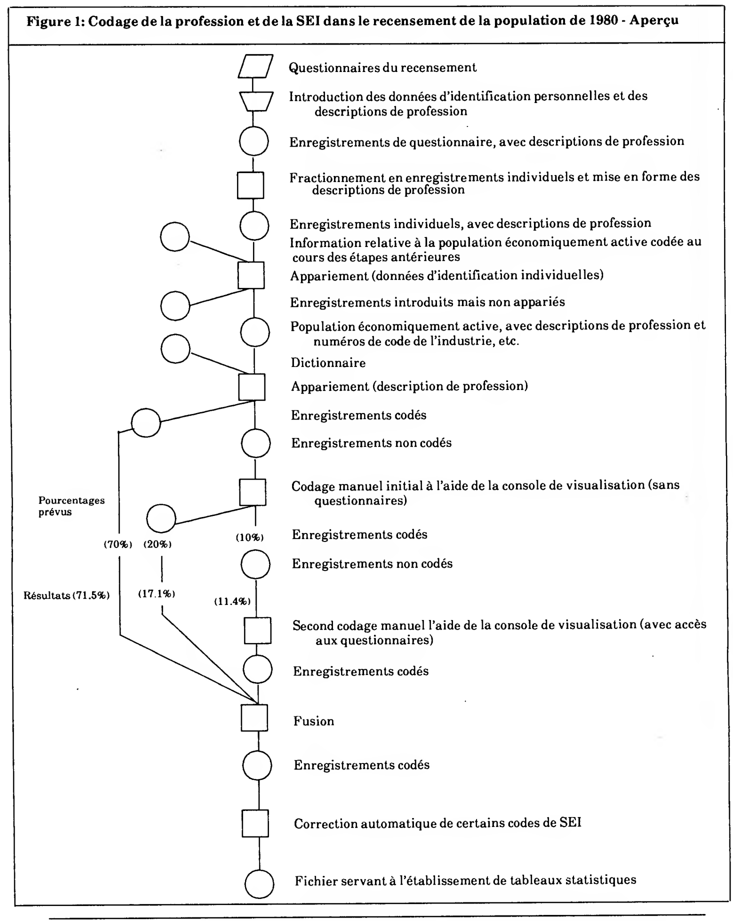
Nous allons aborder maintenant le codage de la profession puisque le système a été conçu, à l'origine, pour ce genre de codage. Le codage de la SEI a été ajouté plus tard et le système ne convient pas "parfaitement" au codage de cette variable.

12.2 Aperçu du système de codage

On trouve à la figure 1 l'organigramme du système de codage.

Les descriptions de profession et les données d'identification personnelles qui figurent sur les questionnaires du recensement sont d'abord introduites en mémoire. On appelle enregistrement de questionnaire l'information introduite à partir d'un questionnaire. Un enregistrement de questionnaire peut contenir un ou deux enregistrements individuels. Après l'introduction en mémoire, les enregistrements de questionnaire sont répartis en enregistrements individuels et les descriptions de profession introduites en mémoire sont en même temps mises en forme.

Figure 1: Codage de la profession et de la SEI dans le recensement de la population de 1980 - Aperçu



Dans le processus de mise en forme, les signes spéciaux (points, lignes, etc.) et les préfixes (1er, vice, etc.) sont supprimés et les autres éléments de la description de profession sont groupés les uns à la suite des autres.

Le fichier ainsi constitué est comparé à un fichier sur la population économiquement active qui a été recensée. Il arrive que certains enregistrements ne puissent être appariés si, par exemple, la profession a été introduite en mémoire pour une personne non économiquement active. Ces enregistrements non appariés ne sont pas utilisés par la suite. En revanche, il arrive qu'aucune description de profession ne soit introduite pour des personnes économiquement actives, en raison probablement du fait que la profession n'est pas indiquée sur le questionnaire. Il arrive parfois que la description de profession figure sur le questionnaire, mais soit omise au moment de l'introduction en mémoire.

Il faut évidemment que toutes les personnes économiquement actives soient codées, ou du moins classées dans des catégories "fourre-tout" lorsque la profession est inconnue. Aux fins de l'appariement, les numéros de code relatifs à l'activité, à l'industrie, à la catégorie d'établissement, etc., sont tirés du fichier sur la population économiquement active.

À la suite de l'appariement, nous obtenons un fichier qui renferme, entre autres, les éléments suivants:

- données d'identification personnelles;
- description de profession introduite et mise en forme (sauf les exceptions mentionnées ci-dessus);
- numéro de code de l'industrie;
- numéro de code de la catégorie d'établissement;
- taille de l'établissement.

Ce fichier est trié en fonction des descriptions de profession mises en forme et des numéros de code de l'industrie, puis comparé au contenu du dictionnaire stocké en mémoire. Si une description de profession mise en forme est trouvée dans le dictionnaire, la profession et la SEI sont codées.

Le dictionnaire des professions utilisé pour le recensement se compose d'un PLEX et d'un SLEX. On en trouvera une description plus détaillée à la page suivante.

Le codage manuel s'effectue à l'aide d'une console de visualisation et comprend deux étapes. Le codage manuel initial est exécuté sans accès aux questionnaires. Les enregistrements qui ne peuvent être codés sont laissés "en blanc" pour être codés plus tard, à la seconde étape de codage manuel. Cette seconde étape est exécutée à l'aide des questionnaires. Puis, les enregistrements codés automatiquement et les enregistrements codés au cours des premières et secondes étapes sont groupés en un seul fichier.

Enfin, certains numéros de code de SEI sont corrigés automatiquement à l'aide d'une question précise du questionnaire, pour laquelle la variable a été codée au cours d'une étape antérieure. Cette question permet de savoir si le répondant est un employeur ou un employé, information qui représente un aspect important du code de SEI.

12.3 Le dictionnaire

Le dictionnaire comporte un PLEX et un SLEX. Un extrait du PLEX apparaît au haut de la page suivante.

Pour qu'il y ait appariement, il doit y avoir concordance exacte entre la description de profession introduite, y compris toute information connexe, et une description contenue dans le dictionnaire PLEX. Le PLEX utilise, comme information connexe, l'industrie (NARG), la catégorie d'établissement (SEKT) et la taille de l'établissement (S).

Comme le codage a été effectué en deux étapes, la situation se prêtait particulièrement bien au codage automatisé. Le genre d'activité, l'industrie et d'autres variables ont d'abord été codés manuellement. Puis, la profession et la SEI ont été codées automatiquement. Comme nous l'avons déjà souligné, ceci permet d'épargner du temps au moment de publier les résultats relatifs aux variables codées lors de la première étape. De plus, le codage en deux étapes permet d'utiliser l'information connexe pendant

ALFABETISKT PRIMÄRLEXIKON

YKOD	SEI		NÄRG	SEKT	S
793	11	TAKARBETARE			
793	11	TAKLÄGGARE			
793	11	TAKMONTÖR			
032	60	TANDL			
032	60	TANDLÄKARE			
044	36	TANDSKÖT			
044	36	TANDSKÖTERSKA			
044	36	TANDSKÖTERSKEELEV			
044	36	TANDSKÖTERSKEPRAKTIKANT			
744	46	TANDTEKNIKER			
032	60	TANDVLÄKARE			
044	12	TANDVÅRDSBITRÄDE			
633	12	TANKBILSCHAFFÖR			
801	11	TAPETFABRIKSARBETARE			
781	21	TAPETSERARE	50		
714	21	TAPETSERARE	*		
781	21	TAPETSÖR	50		
714	21	TAPETSÖR	*		
801	21	TAPETTRYCKARE			
504	11	TAPPARE	2		
821	11	TAPPARE	31		
772	11	TAPPARE	33		
736	11	TAPPARE	37204		
731	11	TAPPARE	37		

le processus de codage automatisé. Nous pensons que les excellents résultats (présentés ci-dessous) du codage automatisé lors du recensement de 1980 sont, dans une large mesure, dus au fait que l'information connexe a été utilisée.

Si l'on examine le listage ci-dessus, on voit, par exemple, que la description "TAKLÄGGARE" (couvreur) reçoit toujours les numéros de code 793 et 11 pour YKOD et SEI, respectivement. La description "TAPETSERARE" (rembourreur) reçoit les numéros de code 781 et 21, respectivement, si le numéro de code de l'industrie est 50 (construction). Pour tous les autres numéros de code de

ADJUNK	052	56	ADJUNKTBIMAKEHÖGSTADIET	93101
ADJUNK	052	56	ADJUNKTBIOLOGIKEMI	93101
ADJUNK	052	56	ADJUNKTBIOLOGIKEMI	93101
ADJUNK	052	56	ADJUNKTBIOMATEMATIK	93101
ADJUNK	052	56	ADJUNKTEKOÄMNEN	93102
ADJUNK	052	56	ADJUNKTENG	93102
ADJUNK	052	56	ADJUNKTENGELSKAFRANSKA	93102
ADJUNK	052	56	ADJUNKTENGELSKAOCHTYSKA	93101
ADJUNK	052	56	ADJUNKTENGELSKATYSKA	93101
ADJUNK	052	56	ADJUNKTENGELSKATYSKA	93102
ADJUNK	052	56	ADJUNKTENGELSKATYSKAMETEMAT	93102
ADJUNK	052	56	ADJUNKTFILOSOFIOCHMATEMAT	93102
ADJUNK	052	56	ADJUNKTFYSIKOMATEMATIK	93102
ADJUNK	052	56	ADJUNKTFÖRETAGSEKONOMI	93102
ADJUNK	052	56	ADJUNKTGRUNDSKOLANSHÖGSTADIU	93101
ADJUNK	052	56	ADJUNKTGYMNASIE	93102
ADJUNK	052	56	ADJUNKTGYMNASIESKOLANKOMVUX	93102
ADJUNK	052	56	ADJUNKTGYMNASIET	93101
ADJUNK	052	56	ADJUNKTHISTORIASVENSKARELIO	93102
ADJUNK	052	56	ADJUNKTHÖGST	93101

STATISTISKA CENTRALBYRÅN 1982-08-19

FOLK-OCH BOSTADSRÄKNINGEN 1980

YRKEN SOM EJ MATCHAT MOT PLEX MED FREKVENS STÖRRE ÄN 1

KOMPRIMERAD YRKESBESKRIVNING	ANTAL
BYGGTRÄ	002
BYRÅDIREKTÖRNATURVÅRD	002
BYRÅINTENDENT	002
BYRÅSEKRFÖRSÄLJNING	002
BÅTTRAFIKÄGARE	002
BÄDDBITRÄDE	002
CEMENTKVARNSOPERATÖR	002
CHARGERARE	002
CHARKARB	002
CHARKARBETARE	003
CHARKFÖRESTÅNDARE	003
CHARKUTERIBITR	003
CHARKUTERISTSTYCKARE	002
CHARKUTERISTTILLVERKNING	002

l'industrie (=*), la description "TAPETSERARE" reçoit les numéros de code 714 et 21, respectivement.

Le SLEX renferme des parties de mots du genre "ADJUNK" (qui est une partie du mot ADJUNKT qui signifie quelque chose comme "professeur adjoint d'école secondaire"). L'objectif visé, c'est qu'une telle partie de mot convienne à de nombreuses variantes d'une description de profession. "ADJUNK", par exemple, convient à toutes les variantes tel qu'il est exposé au bas de la page précédente.

Il peut très facilement arriver qu'un mot SLEX convienne à la "mauvaise" description de profession. Il est difficile d'éviter ce genre d'erreur au moment de la création du SLEX. Pour réduire les erreurs de codage dues à un SLEX "approximatif", on peut, entre autres, utiliser l'information connexe, comme les numéros de code d'industrie. Dans l'exemple au bas de la page précédente, "ADJUNK", pour le numéro de code d'industrie 931 (instruction), est codé à l'aide des numéros de code 052 et 56, respectivement.

D'après notre expérience, un dictionnaire de descriptions de profession SLEX, sans information connexe, donne beaucoup trop d'erreurs de codage. En revanche, nous pensons qu'il est possible de constituer un SLEX efficace si l'on utilise des mots de longueurs différentes et d'autres renseignements, en plus de l'information relative à l'industrie.

La Suède est divisée en 24 comtés et, pour le recensement, le codage s'effectue par comté. Lorsque l'appariement est terminé pour un comté, deux listes sont établies.

La première liste est la liste de fréquences, dont un extrait est presenté ci-dessus.

Cette liste contient les descriptions de profession que le dictionnaire n'a pas réussi à coder et qui reviennent au moins deux fois dans le fichier d'entrée.

Dans l'exemple ci-dessus, on voit, par exemple, que trois enregistrements contenant la description de profession "CHARKFÖRESTÅNDARE" (gérant de boucherie) n'ont pas été codés.

Lorsque le codage relatif à un comté est terminé, la liste de fréquences est lue et de nouvelles descriptions de profession sont introduites dans le PLEX. De plus, des listes de contrôle fournissent l'information connexe qui sert à corriger le PLEX. La taille du PLEX est passée de 4,000 à 11,000 enregistrements au cours de la phase d'exploitation.

L'autre liste, la liste SLEX, concerne les descriptions de profession qui ont été codées par le SLEX. Un extrait en est présenté à la page suivante.

FOLK-OCH BOSTADSRÄKNINGEN 1980 DATUM 1982-10-06 SIDA 76

POSTER SOM MATCHAT MOT SLEX

ORDDEL	YRK	SEI	REDYRKE	NÄRG
ÖVERLÄ	031	57	ÖVERLÄKAREKIRKLIN	93310
ÖVERLÄ	031	57	ÖVERLÄKAREKIRURGI	93310
ÖVERLÄ	031	57	ÖVERLÄKAREKIRURGKLINIKEN	93310
ÖVERLÄ	031	57	ÖVERLÄKAREKLINIKCHEFLÅNGVÅRD	93310
ÖVERLÄ	031	57	ÖVERLÄKAREKLINIKERCHEF	93310
ÖVERLÄ	031	57	ÖVERLÄKAREMEDKLIN	93310
ÖVERLÄ	031	57	ÖVERLÄKARERÖNTGEN	93310
ÖVERLÄ	031	57	ÖVERLÄKARERÖNTGEN	93310
ÖVERLÄ	031	57	ÖVERLÄKGYNEKOLOGI	93310

Dans la liste SLEX, on peut facilement repérer les erreurs de codage dues au caractère approximatif du SLEX.

La taille du SLEX n'a pas augmenté autant que celle du PLEX, car nous n'avons pas eu le temps de trouver et d'essayer de nouveaux mots SLEX. Le SLEX renferme un peu plus de 500 mots. Comme nous l'avons souligné précédemment, nous pensons qu'il serait possible de créer un SLEX beaucoup plus efficace si on utilisait l'information connexe.

Pour toute la phase d'exploitation, le degré de codage a été de 71.5%, dont environ 68% par le PLEX et 3% par le SLEX. Le degré de codage a varié de 67.2% à 76.6% selon le comté. Comme nous visions un objectif global de 70%, nous avons réussi au-delà de nos espérances.

Le coût de l'exécution du programme d'appariement est négligeable. Examinons l'exemple ci-dessous: les descriptions relatives à un comté renfermant 341,529 personnes économiquement actives ont été comparées au contenu d'un PLEX de 10,291 descriptions et à celui d'un SLEX de 513 mots. Les résultats ont été les suivants:

	Nombre d'enregistrements codés	Degré de codage (%)
PLEX	246,652	72.2
SLEX	8,339	2.2
Total	**254,991**	**74.7**

Le coût de l'appariement et du codage automatisé a été de 303 couronnes suédoises, soit environ 50 dollars canadiens.

Il convient aussi de souligner que, d'après notre expérience en matière de recensements, il faudrait demander au personnel préposé à l'introduction en mémoire d'introduire exactement l'information qui figure sur les questionnaires (sans dépasser le nombre maximal de caractères préalablement spécifiés, soit 30 en l'occurrence). Nous pensons que ceci permettrait d'obtenir la meilleure combinaison possible de taux et de qualité d'introduction en mémoire.

12.4 Codage manuel initial

Après l'appariement à l'aide du dictionnaire, il reste près de 30% de la population économiquement active à coder. Il faut coder cette proportion manuellement. Dans le cadre d'un recensement, ce codage s'effectue en deux étapes. Le codage manuel initial est exécuté au moyen des consoles de visualisation sans accès aux questionnaires. Vingt enregistrements apparaissent en même temps sur la console de visualisation. Pour chaque personne, l'image affichée sur l'écran indique la description de profession, le numéro de code de l'industrie, la catégorie d'établissement, la taille de l'établissement et le genre d'activité.

Les codeurs utilisent une liste alphabétique de professions, qui contient plus de 12,000 descriptions de profession officielles, ainsi que les numéros de code de profession et de SEI correspondants. La principale règle à respecter est la suivante: le codeur doit trouver, dans la

liste de professions, "exactement" la même profession que celle qui est indiquée sur l'écran. Après avoir trouvé la profession dans la liste, le codeur introduit par l'intermédiaire du clavier les numéros de code correspondants. Les descriptions de profession qui ne peuvent être codées le seront au moment du second codage manuel.

Nous avions prédit que 20% du nombre total des enregistrements seraient codés lors du codage manuel initial. Nous avons obtenu un pourcentage de 17.1%. Les codeurs qui ont exécuté le codage manuel initial ont réussi à coder, en moyenne, 217 enregistrements l'heure (ce chiffre comprend les enregistrements non codés).

12.5 Second codage manuel

Le second codage manuel permet de coder le reste des enregistrements. Ce codage s'effectue par l'intermédiaire de consoles de visualisation et à l'aide des questionnaires. Nous avions prévu qu'il resterait 10% des enregistrements à ce stade, il en restait 11.4%.

Le second codage manuel est très fastidieux. En réalité, cette étape s'apparente beaucoup au codage classique d'environ 10% des descriptions posant le plus de difficultés. En moyenne, les codeurs préposés au second codage manuel ont réussi à coder 27 enregistrements l'heure.

Comme nous le voyons, cette subdivision en deux étapes rend le processus de codage beaucoup plus facile. Pendant la première étape, on peut maintenir la rapidité de codage à un niveau très élevé puisque le codage s'effectue sans accès aux questionnaires.

12.6 Quelques remarques d'ordre général

Comme nous l'avons déjà mentionné, le degré de codage de la profession et de la SEI a été de 71.5% lors du recensement de la population de 1980. Les calculs effectués avant de prendre la décision de recourir au codage automatisé avaient révélé qu'un degré de codage de 60% serait rentable.

Il convient aussi de souligner que la grande rapidité de codage obtenue à l'étape du codage manuel initial est due, en grande partie, au fait que les descriptions de profession sont déjà introduites dans l'ordinateur, de sorte que le codage peut s'effectuer sans consultation des questionnaires. Ceci permet d'épargner beaucoup de temps.

Nous ne savons évidemment pas ce qu'aurait coûté un système imaginaire de codage manuel classique de la profession et de la SEI lors de ce recensement, mais nous avons lieu de croire que le codage automatisé nous a permis d'économiser au moins un million de couronnes suédoises (soit environ 170,000 dollars canadiens), c'est-à-dire environ 10% du coût global du codage de la profession et de la SEI.

Nous n'avons cependant pas eu recours au codage automatisé uniquement pour des raisons d'argent. Il aurait, en effet, été impossible, pour Statistique Suède, d'obtenir le personnel nécessaire à l'exécution du codage dans un délai raisonnable. Le codage automatisé a fait passer de 4,000,000 à 1,200,000 le nombre de descriptions à coder manuellement au point de vue de la profession et de la SEI et a rendu possible l'utilisation de deux systèmes de codage manuel.

Nous pensons aussi que l'introduction des descriptions de profession dans l'ordinateur présente en soi de grands avantages. Tout comme les descriptions d'achat, les descriptions de profession renferment plus de renseignements qu'un numéro de code unique. Ce "surplus" d'information peut se révéler utile pour les recherches médicales futures, par exemple.

Malheureusement, aucune évaluation de la qualité de codage n'a été faite. Toutefois, nous savons que le PLEX permet de coder presque sans erreur près de 70% de la population économiquement active. Les listes SLEX ont été vérifiées soigneusement pendant toute la phase d'exploitation et, compte tenu du fait que le SLEX n'a été utilisé que dans une très faible proportion des cas, on peut supposer que son inexactitude relative n'influe pas sérieusement sur le taux global d'erreurs. En outre, après le codage relatif à chaque comté, nous avons vérifié les listes de contrôle, c'est-à-dire que nous avons fait sortir un échantillon des enregistrements et

vérifié les numéros de code pour le comté. Cette méthode a permis d'améliorer continuellement le dictionnaire et les instructions de codage. Les listes nous ont donné des estimations approximatives des taux d'erreur et nous croyons fermement que le taux d'erreur relatif à la profession est plus faible, dans ce recensement, que le taux estimatif de 8% obtenu lors du recensement de 1975.

13. Autres applications

13.1 Codage de la profession, de la classification socio-économique (SEI) et de l'affiliation syndicale dans Survey of Living Conditions (SLC)

Dans le cadre de la SLC en cours, toute l'information numérique et une partie de l'information verbale sont introduites en mémoire afin que la mise en forme soit plus efficace. L'information verbale introduite peut accessoirement servir au codage automatisé. C'est le cas des variables "profession" et "SEI". Le fichier ainsi constitué est mis en forme et les descriptions de profession sont comparées au contenu d'un dictionnaire PLEX. S'il y a appariement, les numéros de code de la profession et de la SEI sont listés ensemble, ainsi que toute autre information perforée à partir des questionnaires. Puis, le personnel préposé au codage vérifie les numéros de code attribués automatiquement et les modifie au besoin. De plus, les descriptions non codées sont codées sur la liste. Le PLEX utilisé pour la Survey of Living Conditions est une partie du PLEX utilisé lors du recensement de 1980. Il s'agit de la partie qui ne renferme pas d'information connexe.

Le système "semi-automatisé" fonctionne bien et on l'a récemment élargi pour y ajouter la variable "affiliation syndicale" (degré de codage de 60%). On envisage de l'élargir encore davantage pour y inclure la variable "instruction".

13.2 Codage de la profession dans les enquêtes sur les étudiants

Statistique Suède exécute des enquêtes permanentes sur les différents groupes d'étudiants. Ces enquêtes sont effectuées quelques temps après que les étudiants ont terminé leurs études. Elles visent à obtenir des renseignements sur leur emploi actuel et sur leurs projets d'avenir.

Presque toute l'information obtenue au moyen des questionnaires remplis au cours de ces enquêtes est introduite en mémoire pour des raisons différentes. Lors de deux enquêtes récentes, l'information introduite a servi au codage automatisé de la profession.

Le PLEX utilisé pour les enquêtes sur les étudiants est une partie du PLEX utilisé lors du recensement de 1980, à savoir la partie ne renfermant pas d'information connexe.

Le degré de codage obtenu se situe autour de 50%. Le reste des descriptions de profession introduites en mémoire peuvent être en grande partie codées manuellement, sans consultation des questionnaires. Ce codage sans accès aux questionnaires s'effectue beaucoup plus rapidement que le codage classique. Nous avons eu une expérience analogue lors du recensement de 1980; les écarts relatifs à la rapidité de codage, entre le premier et le second codage manuel, sont considérables. C'est là, à notre avis, un avantage souvent négligé du codage automatisé. En effet, la rapidité de codage de la partie manuelle d'un système de codage automatisé peut être sensiblement plus grande que celle d'un système de codage classique, c'est-à-dire entièrement manuel.

L'utilisation du codage automatisé lors des enquêtes sur les étudiants a été fructueuse. Pour ces enquêtes, on utilise maintenant le système de codage automatisé de la profession.

13.3 Prêts de livres

Le Swedish Author's Fund verse des sommes aux auteurs en proportion de la popularité de leurs livres auprès des abonnés des bibliothèques publiques de la Suède. Cette gratification basée sur des échantillons prélevés dans diverses bibliothèques est versée une fois par année. L'enquête est exécutée par Statistique Suède moyennant une commission.

Le contexte général de traitement informatisé des données, selon lequel

une liste alphabétique des noms d'auteur et des titres de livres introduits en mémoire est générée, convient particulièrement bien au codage automatisé. Dans ce genre d'application, un système automatisé peut facilement concurrencer un système manuel. Un degré de codage même faible rend le système automatisé rentable puisque l'introduction en mémoire est "gratuite". La seule condition, c'est que les coûts informatiques, par enregistrement, soient inférieurs au coût du codage manuel.

On n'utilise qu'un dictionnaire PLEX ayant un taux de descriptions non ambiguës de 100% puisque chaque erreur peut influer sensiblement sur les gratifications remises.

Le système est en exploitation depuis 1978. Pendant cette période, le nombre de noms d'auteur et de titres de livres contenus dans le dictionnaire est passé de 6,900 à 65,000. Pendant ce laps de temps, le degré de codage a augmenté, passant de 33% à 80%. Des études d'évaluation menées il y a quelques années ont révélé que le dictionnaire renfermant 65,000 descriptions n'était pas exact à 100%. On a donc constitué un dictionnaire révisé de 35,000 descriptions. Le degré de codage a chuté, passant de 80% à 68%. Comme le système était rentable dès le départ, il est maintenant extrêmement profitable.

Perspectives d'avenir du codage automatisé

Au moment de la conception d'une activité de codage, il est évident qu'on peut envisager le codage automatisé de certaines variables. Nous pensons cependant que le succès du codage automatisé dépend de la complexité de la langue. Il semble que, sous ce rapport, le suédois soit plus simple que l'anglais.

Pour la plupart de nos expériences et de nos applications, nous avons adopté des méthodes très simples. Tous les travaux pour lesquels nous avons utilisé des méthodes complexes n'ont été ni très fructueux, ni très étendus. Le fait que des degrés de codage relativement faibles (de 65% à 70%) se soient révélés rentables, a probablement nui à l'élaboration de méthodes plus efficaces. Nous devrions tendre vers des systèmes encore plus profitables; il faudrait que le degré de codage augmente de 10 à 15 points en ce qui concerne, par exemple, le codage des biens ou de la profession. Pour ce faire, il faudrait adopter des méthodes plus complexes, mais aussi modifier le code sous certains rapports. La fusion de catégories différentes est parfois interdite en raison d'obligations envers les utilisateurs des données. Il n'est peut-être pas trop déraisonnable d'apporter des modifications aux codes afin de réduire le coût du codage. Cette option mérite certainement d'être envisagée plus souvent en temps de pénurie des ressources financières.

Il est également possible d'améliorer le niveau de codage en stockant l'information connexe dans les dictionnaires et en utilisant des dictionnaires SLEX plus efficaces.

Le codage automatisé est bien implanté. En raison des lois régissant le marché du travail en Suède, il est difficile d'engager des codeurs pour des travaux temporaires, comme pour le codage à l'occasion d'un recensement. Nous devons compter sur notre personnel permanent, et le codage automatisé nous dépanne en périodes de pointe.

Jusqu'ici, notre stratégie a consisté à tester les variables les plus faciles d'abord. Nous devons maintenant nous attaquer aux variables plus difficiles et rendre les dictionnaires et les procédures de soutien plus efficaces.

BIBLIOGRAPHIE

Appel, M.V. et Hellerman, E., 1983. "Census Bureau Experience with Automated Industry and Occupation Coding", American Statistical Association, **Proceedings of the Section on Survey Research Methods**, p. 32-40.

Appel, M.V. et Scopp, T., 1985. "Automated Industry and Occupation Coding", Paper presented to the Census Advisory Committee of the American Statistical Association and on Population Statistics at the Joint Advisory Committee Meeting, 25 avril 1985, Rosslyn, Virginia.

Bailar, B.A. et Dalenius, T., 1969. "Estimating the Response Variance Components of the U.S. Bureau of the Census' Survey Model", **Sankhya**, série B, vol. 31, parties 3 et 4, p. 341-360.

Bäcklund, S., 1978. "Automatisk kodning. Beskrivning av programvara och programvaruhantering", Note de service, Statistique Suède (en suédois).

Corbett, J.P., 1972. "Encoding from Free Word Descriptions", Note de service, U.S. Bureau of the Census.

Dalenius, T. et Lyberg, L., s.d. "An Experimental Comparison of Dependent and Independent Verification of Coding", Note de Tore Dalenius à Leon Pritzker.

Harvig, H., 1973a. "Kontrollkodning av dodsbevis", Note de service, Statistique Suède (en suédois).

Harvig, H., 1973b. "Kontrollkodningsexperiment på blanketter för inskrivningsuppgifter till högre studier", Note de service, Statistique Suède (en suédois).

Knaus, R., s.d., "Syntactically Based Classification from Natural Language Responses", Note de service, U.S. Bureau of the Census.

Knaus, R., 1978a. "Inference by Semantic Pattern Matching in Industry Classification", Note de service, U.S. Bureau of the Census.

Knaus, R., 1978b. "Automated Industry Coding - An Artificial Intelligence Approach", Note de service, U.S. Bureau of the Census.

Knaus, R., 1979. "A Similarity Measure on Semantic Network Nodes", Paper presented at the Classification Society Annual Meeting, Gainesville, Florida, 1979.

Knaus, R., 1983. "Methods and Problems in Coding Natural Language Survey Data", American Statistical Association, **Proceedings of the Section on Survey Research Methods**, p. 51-60.

Lakatos, E., 1977a. "Automated I & O Coding", Note de service, U.S. Bureau of the Census.

Lakatos, E., 1977b. "Computerized Coding of Free Verbal Responses", Note de service, U.S. Bureau of the Census.

Lyberg, L., s.d. "Beroende och oberoende kontroll av kodning", Rapport nº 4, Forskningsprojektet FELI UNDERSÖKNINGAR, Stockholms universitet, Stockholm (en suédois).

Lyberg, L., 1981. **Control of the Coding Operation in Statistical Investigations - Some Contributions,** Thèse de Ph.D., Urval nº 13, Statistique Suède.

Lyberg, L., 1983. "The Development of Procedures for Industry and Occupation Coding at Statistics Sweden", **Statistical Review**, p. 139-156.

Lyberg, L., Nordling, P. et Elmdahl, J., 1973. "Kodningskvaliteten i lararregistret", Note de service, Statistique Suède (en suédois).

Minton, G., 1969. "Inspection and Correction Error in Data Processing", **Journal of the American Statistical Association**, vol. 64, p. 1256-1275.

Olofsson, A., 1976. "Kvalitetskontroll av naringsgrenskodningen i AKU hosten - 74", Note de service, Statistique Suède (en suédois).

O'Reagan, R.T., 1972. "Computer-Assigned Codes from Verbal Responses", **Communications from the ACM**, vol. 15, nº 6, p. 455-459.

Owens, B., 1975. "The Corbett Algorithm for Coding from Free Word Descriptions", Note de service, U.S. Bureau of the Census.

L'AVENIR DES LOGICIELS GÉNÉRAUX

OU - LE LOGICIEL S'USE-T-IL?

MIKE JEAYS

DIVISION DES SERVICES ET DU DÉVELOPPEMENT INFORMATIQUES
STATISTIQUE CANADA

Renseignements généraux

Dans le recensement, le traitement consiste en quatre étapes principales que voici:

1. Saisie des données
2. Contrôle et imputation
3. Totalisation (production de tableaux)
4. Analyse des données

Un système élaboré de gestion et de contrôle sert à faire en sorte que le traitement soit exécuté selon le calendrier établi et que tous les questionnaires soient bien pris en compte.

Le logiciel général joue un rôle important dans les recensements de la population qu'effectue Statistique Canada depuis 1971. À cette époque, le système de génération de tableaux STATPAK était élaboré par le Bureau. La principale caractéristique de ce système est le langage appelé TARELA qui peut servir à spécifier un vaste éventail de tableaux à partir des fichiers de recensement. (Les fichiers de recensement consistent essentiellement en une collection de trois relations: secteurs de dénombrement, ménages et particuliers.) STATPAK, spécialement dans ses premières versions, s'articule autour de la structure des fichiers de recensement et utilise beaucoup d'instructions qui renvoient expressément aux liaisons entre ces fichiers. Il était employé avec un progiciel appelé CASPER afin de produire des tableaux pour publication.

Le succès de cette méthode a incité le Bureau à se lancer dans la réalisation d'autres logiciels généraux pour le recensement de 1976. On a ensuite abordé les problèmes liés à la gestion d'une base de données volumineuse. On s'est vite aperçu qu'aucun des SGBD disponibles sur le marché ne pouvait traiter des fichiers de 24 millions d'enregistrements couvrant plusieurs volumes de disques tout en offrant des temps de réponse raisonnables et en assurant une utilisation rationnelle de la mémoire à disques. La solution était de développer le système RAPID qui utilise une architecture de fichiers transposés pour optimiser la performance au moment de la production des tableaux. (Le gros du traitement du recensement consiste à produire un vaste éventail de tableaux. Les variables sont soumises une à la fois au contrôle et à l'imputation, d'où l'utilité d'une architecture de fichiers transposés.) Par cette exigence, ce système diffère sensiblement de la plupart des systèmes commerciaux qui sont optimisés pour la recherche par enregistrement et qui offrent des fonctions de mise à jour simultanée. L'architecture de fichiers transposés, où toutes les valeurs d'une variable sont mémorisées consécutivement, permet d'atteindre le niveau élevé de rendement recherché. Son utilisation a donné d'excellents résultats et RAPID peut encore dépasser tous les logiciels commerciaux qui exécutent ce type de traitement. Aucun changement sous ce rapport n'est prévu avant 1991.

Le contrôle et l'imputation des variables qualitatives (celles qui comportent un nombre relativement petit de valeurs discrètes) étaient exécutés par un second progiciel général. Ce dernier, appelé CANEDIT, a été créé à temps pour le recensement de 1976. Il s'inspirait d'une méthode décrite dans un article de Fellegi et Holt. Les contrôles sont spécifiés comme des instructions "CONTRADICTOIRES". Ils consistent en un ensemble de relations logiques entre les variables d'un enregistrement logique jugées impossibles. Un exemple évident serait celui d'une personne qui serait mariée, mais qui n'aurait que 3 ans. L'algorithme détermine le nombre minimal de zones à modifier pour que l'enregistrement satisfasse à toutes les conditions de contrôle, puis insère dans les zones invalides les valeurs correspondantes d'un enregistrement "similaire" qu'il sait exempt d'erreur. Ce processus s'appelle l'imputation. Au recensement de 1981, le nombre de variables recueillies était considérablement plus élevé, étant passé de 15 à 46. Cette augmentation a rendu difficile sinon impossible l'utilisation de CANEDIT pour exécuter la totalité du contrôle et de l'imputation et un nouveau progiciel général appelé SPIDER a été élaboré. Ce progiciel accepte des contrôles sous forme de tableaux de décision. Après des analyses de cohérence, il génère des instructions en langage PL/I qui sont ensuite compilées en un programme qui lui-même exécutera les contrôles et imputations spécifiés dans les tableaux de décision originaux. Bien que SPIDER soit de nature essentiellement générale, un certain nombre d'extensions ont été écrites spécialement

pour le traitement du recensement et ce n'est que plus tard qu'une version partiellement réduite a été mise au point pour servir à d'autres projets. CANEDIT et SPIDER ont été utilisés dans le recensement de 1981 et le seront de nouveau dans celui de 1986. SPIDER peut faire tout ce que CANEDIT fait, mais la forme des spécifications à l'entrée qu'il accepte diffère complètement. Les deux progiciels ensemble ont éliminé la nécessité d'écrire des centaines de milliers de lignes d'instructions en langages de programmation courants. Les coûts d'écriture et de maintenance de ces instructions auraient été beaucoup plus élevés que ceux qu'a occasionné l'utilisation du logiciel.

Une caractéristique importante des deux progiciels est qu'ils fournissent de l'information en retour sur les spécifications à l'entrée avant d'entreprendre tout travail. Par exemple, CANEDIT génère un ensemble de "contrôles implicites" à partir de ceux qui ont été spécifiés. Ceux-ci peuvent être vérifiés; un contrôle implicite qui est de toute évidence erroné indique à l'utilisateur que les contrôles qu'il a spécifiés doivent contenir une erreur. Ces derniers peuvent être corrigés à volonté.

Le traitement du recensement de 1986 se fera autant que possible au moyen du même logiciel qu'en 1981. La décision de ne pas élaborer de nouveaux logiciels a déjà permis de réaliser des économies considérables et devrait contribuer au déroulement sans heurts du recensement de 1986. L'investissement s'est avéré rentable.

Problèmes et risques éventuels du recensement de 1991

Non, les logiciels ne s'usent pas!

Les logiciels, contrairement à bon nombre d'autres biens, ne s'usent pas. Cependant, pour un éventail de raisons, les logiciels semblent tous nécessiter une maintenance régulière. Le cadre matériel et logiciel dans lequel ils sont utilisés risque généralement de subir un certain nombre de changements additionnels. Ces changements souvent sans grande portée sont considérés comme étant "à compatibilité ascendante" - c.-à-d. sans incidence sur les applications existantes. Toutefois, les effets cumulés de ces changements peuvent parfois être plus sérieux et un logiciel qui n'a pas servi depuis un certain temps peut être tout à fait inutilisable lorsqu'on en a besoin. Les changements nécessaires pour surmonter les difficultés peuvent être mineurs, mais si personne d'expérimenté n'est disponible à bref délai, les conséquences sur un calendrier de production serré peuvent être catastrophiques. La clé du succès est une utilisation continue et une vérification approfondie des logiciels peu de temps avant le traitement d'une application d'importance.

À part les changements apportés au système d'exploitation sous la commande duquel une application ou un progiciel général est exécuté, il y a les changements que l'on fait subir au matériel lui-même. Les ordinateurs actuels sont très différents de ceux de 1971. Beaucoup plus puissants, ils ont en outre un jeu d'instructions et une mémoire virtuelle supérieurs, des systèmes d'entrée-sortie plus sophistiqués et des unités de mémoire à accès direct de loin plus efficaces.

La fonction principale du système d'exploitation est d'isoler les applications de la plupart de ces changements. Les services assurés par le système d'exploitation ont augmenté avec le temps, mais les caractéristiques qu'il avait en 1971 sont toujours là. Nous pouvons sans crainte de nous tromper prédire que ce sera aussi le cas en 1991 et presque certainement longtemps après. MVS (ou un système d'exploitation qui offre des services compatibles) sera certainement toujours utilisé d'ici la fin du siècle; l'investissement fait dans le monde entier en logiciels exécutables sous la commande de ce système est beaucoup trop important pour que l'on puisse croire le contraire.

Les innovations réalisées dans les unités de mémoire à accès direct nécessitent le plus l'adaptation continue du logiciel et ces appareils vont certainement continuer à augmenter de performance et de capacité très longtemps encore. Les unités de lecture seulement (les disques optiques) seront vraisemblablement disponibles très bientôt et devraient convenir fort bien à un système d'extraction de données du recensement. Ces nouveaux appareils seront très certainement exploitables sous la commande de MVS avec compatibilité ascendante. Le mécanisme de l'offre et de la demande fera en sorte que l'adaptation aux nouvelles unités de mémoire se fera sans difficultés majeures, à condition toutefois que le logiciel soit écrit de telle façon qu'il puisse s'adapter facilement aux changements des paramètres, comme par exemple les longueurs des pistes.

Problèmes du soutien des logiciels généraux

Malgré les commentaires ci-dessus, il est clairement nécessaire que les groupes d'élaboration

des logiciels puissent apporter tous les changements et toutes les améliorations qui s'imposent. Ces changements peuvent être classés en trois catégories:

1. Changements nécessités par des modifications sans compatibilité ascendante au cadre d'exploitation: nouveau matériel, changements au système d'exploitation, etc. Ces changements sont clairement indispensables si l'on veut continuer à utiliser le logiciel.

2. Corrections d'erreurs reconnues dans le logiciel (mises au point). De nouvelles erreurs ne paraissent pas spontanément dans le logiciel - elles ont soit été introduites incidemment avec une autre mise au point ou elles étaient déjà là mais n'avaient pas été relevées. À l'exception du cas rare où des résultats erronés ont été obtenus dans le passé mais n'ont pas été remarqués, le logiciel peut continuer à être utilisé pour le sous-ensemble de ses possibilités qui se prêtaient bien à l'exécution.

3. Addition de nouvelles fonctions. Il ne s'agit pas ici de maintenance réelle. Si l'on est souvent tenté de considérer l'addition de fonctions mineures comme de la maintenance, il convient de distinguer clairement les deux opérations.

Pour chaque progiciel dont il doit assurer le soutien, le groupe responsable du logiciel général doit avoir au moins deux employés qui connaissent en détail les mécanismes internes du système. Sans les occuper à plein temps, cette tâche devrait être une partie importante de leur travail. Ces employés doivent être hautement qualifiés et bien motivés. Un des meilleurs moyens de s'assurer de leur compétence et motivation est de leur demander de réaliser des développements additionnels, souvent en ajoutant des fonctions au logiciel, en leur spécifiant bien de doter tout nouveau logiciel conçu d'une compatibilité ascendante avec les versions antérieures. Le meilleur moyen de justifier et de financer cette activité coûteuse est de garder le logiciel continuellement en usage dans l'organisation. Les logiciels qui conviennent aux enquêtes périodiques et aux recensements seront utilisés régulièrement ce qui garantira leur prompte mise au point. Un logiciel ainsi utilisé régulièrement aura plus tendance à survivre dans l'organisation qu'un logiciel inutilisé durant de longs intervalles. Dans ce dernier cas, les experts, occupés ailleurs, peuvent ne pas être sur place au moment où des difficultés sont décelées et leur absence peut augmenter le temps et le coût d'exécution des changements nécessaires.

Le groupe d'élaboration devrait être fortement encouragé d'écrire le logiciel de manière à en simplifier la maintenance. En dehors d'une bonne documentation, cette exigence implique la nécessité d'éviter les "astuces" qui peuvent causer des difficultés plus tard. Il faudrait veiller à paramétrer les chiffres qui sont susceptibles de changer comme les caractéristiques des unités de mémoire à accès direct. L'adoption de ces pratiques dans le système RAPID a déjà assuré son adaptabilité à un environnement qui a changé de façon significative depuis sa première version et cette expérience devrait être appliquée à toute situation qui s'y prête.

Compétence chez les utilisateurs du recensement

Le problème de garder un noyau adéquat d'employés qualifiés se pose également pour les utilisateurs du logiciel du recensement. Il est difficile de former un nouveau groupe d'utilisateurs s'il ne reste plus de personne compétente qui connaisse à fond tous les détails et tous les astuces qu'il est pratiquement impossible de mettre sur papier. On pourrait faire une analyse coût-avantages et comparer ce qu'il en coûterait de garder du personnel spécialisé et expérimenté mais légèrement sous-exploité et ce qu'il en coûterait de former à nouveau le personnel de l'utilisateur. Pour des progiciels en usage constant comme STATPAK, la question ne se pose pas. Mais pour d'autres logiciels comme CANEDIT et SPIDER, qui ne sont pas toujours utilisés, les difficultés sont réelles.

Y a-t-il de meilleures solutions sur le marché?

Les logiciels disponibles sur le marché évoluent à un rythme accéléré et deviennent de plus en plus complexes et puissants. Lorsque le système RAPID a été créé, il n'y avait pas sur le marché un SGBD qui pouvait traiter des fichiers de 24 millions d'enregistrements. Ce n'est plus le cas aujourd'hui. Toutefois, les demandes du marché commercial ne correspondent pas exactement aux besoins du recensement. Les demandes spéciales les plus importantes sont les suivantes:

1. Des opérations de contrôle et d'imputation automatisées à grande échelle, particulières aux bureaux statistiques.

2. Une grande variété de tableaux spéciaux de formats divers.

3. La nécessité de produire ces tableaux sous forme de publications d'excellente présentation. (Ce besoin va diminuer les publications officielles cédant le pas aux publications spéciales qui répondent aux demandes des utilisateurs.)

4. Une efficacité extrême dans le traitement de fichiers de nombreux millions d'enregistrements.

5. Des résultats bilingues.

Certaines de ces exigences peuvent être satisfaites, du moins en partie, par les logiciels commerciaux. TPL (mis au point par le U.S. Bureau of Labor Statistics) et la procédure de génération de tableaux SAS (lui-même construit sur le modèle de TPL) conviennent tous deux des points de vue résultats et souplesse, mais n'offrent pas le soutien adéquat pour des résultats bilingues. Si donc des systèmes d'extraction et de génération de tableaux sont disponibles, il nous faut développer et tenir à jour notre propre logiciel de contrôle et d'imputation.

Devons-nous continuer à adapter le logiciel?

Comme nous l'avons déjà dit, un des meilleurs moyens de garder un logiciel à jour est de le développer et de l'utiliser constamment. Ainsi en a-t-il été de RAPID qui est utilisé par un certain nombre de projets au Bureau et qui est largement diffusé dans d'autres centres de traitement. Un programme de maintenance et de mise au point constantes a donné une collection de nouveaux utilitaires et une meilleure performance.

Il y a très peu de concepteurs de logiciels généraux compétents sur le marché et pour les former, il faut des années d'expérience. Il est donc recommandé de choisir les fonctions que l'on veut développer et de concentrer les ressources essentiellement sur le logiciel de soutien de ces fonctions.

Pour la période 1986-1991, la fonction candidate la plus importante est celle du contrôle et de l'imputation. CANEDIT ne suffit plus aujourd'hui au contrôle nécessaire. (C'est la raison pour laquelle SPIDER a été conçu.) SPIDER, quant à lui, ne fait pas tout ce qu'il faut. Il est donc suggéré de créer, c'est-à-dire de concevoir et d'écrire, à temps pour le recensement de 1991, un nouveau progiciel en combinant les meilleures caractéristiques des deux et en adoptant toute idée originale qui a vu le jour ces quelques dernières années. Le développement de ce logiciel devrait être fait à même le budget du Bureau, avec l'intention de rendre le produit utilisable à beaucoup d'autres applications à part le recensement.

La seconde fonction candidate est celle du système d'extraction. STATPAK date aujourd'hui et, en pratique, de nombreux tableaux sont générés avec le système SAS. (STATPAK est utilisé pour l'extraction initiale à partir des fichiers de recensement, pour des raisons d'efficacité et à cause de son étroite intégration aux fonctions de sélection géographique.) La recherche d'une meilleure interface entre les procédures de génération de tableaux de SAS et les fichiers de recensement est une forte possibilité. D'une façon plus générale, le Bureau devrait surtout chercher à utiliser efficacement des logiciels commerciaux en construisant des interfaces ou même en sous-traitant la mise au point d'extensions qui répondraient à nos besoins spéciaux. Cette façon de procéder serait plus rentable que d'essayer de développer un nouveau système de génération de tableaux.

Nouveaux besoins du recensement

Les nouveaux besoins du recensement doivent bien sûr être étudiés en détail. Les nouvelles options de collecte de données seront étudiées, et le volume des données recueillies par personne et par ménage va probablement augmenter. Nous aurons toujours besoin d'une grande variété de tableaux et nous devrions toujours être à même de produire la majorité de ces tableaux en utilisant un progiciel général.

Les nouvelles options, comme le codage automatique des professions, devraient être examinées. Les méthodes récentes en intelligence artificielle pourraient nous être utiles ici et un bassin de connaissances devrait être créé au Bureau.

On pourrait aussi se pencher sur d'autres moyens de diffusion de l'information, en particulier le recours aux médias électroniques modernes.

La rentabilité doit aussi être étudiée. En général, on peut affirmer que la recherche d'une meilleure performance par la mise au point revient moins cher que la réécriture entière du logiciel.

Examen des logiciels généraux

Le nombre des utilisateurs des logiciels généraux varie beaucoup selon le progiciel. Par exemple, SAS est de loin le progiciel le plus utilisé dans notre centre et intervient dans environ 25% des traitements par lots. Ses utilisateurs sont nombreux, avec des degrés de connaissances et de compétences variés. CANEDIT par contre n'est utilisé que pour le recensement de la population. RAPID est aujourd'hui utilisé dans quelque 15 projets différents.

Une analyse de l'utilisation de chacun des progiciels permettrait de décider beaucoup plus facilement à quels progiciels on devrait donner le plus de soutien et quels autres devraient être enlevés. Il y a quelques difficultés techniques à mesurer l'utilisation de certains progiciels et des solutions devraient être trouvées.

Une analyse des travaux par lots a récemment indiqué qu'environ 45% de ces travaux dépendaient entièrement ou partiellement d'un ou de plusieurs progiciels généraux. Ce chiffre va sans doute augmenter plus tard lorsqu'on aura adopté la politique proposée qui encourage fortement l'utilisation de ces logiciels commerciaux.

Il y aurait certainement lieu de revoir la collection de logiciels à usage général exploités au Bureau. Cette collection est documentée dans le répertoire du logiciel qui a été largement diffusé dans l'organisation. Certains progiciels sont très répandus et jouissent d'un bon soutien; d'autres sont rarement ou pas du tout utilisés et aucune maintenance n'a été faite récemment si ce n'est pour adapter le logiciel à des changements environnementaux lorsque ceux-ci ont eu une incidence qui a nécessité la maintenance. Il est proposé que cette collection soit examinée en détail par les responsables de la maintenance et par un groupe d'utilisateurs et qu'un rapport recommandant les mesures à prendre, comme le désengagement officiel à l'égard du soutien du logiciel inutilisé, soit rédigé. Ce rapport devrait ensuite être largement diffusé parmi les utilisateurs afin de leur donner l'occasion d'exprimer leurs commentaires. Les conclusions finales, qui tiendront compte de l'avis du plus grand nombre d'utilisateurs possibles, devraient être présentées simultanément au Comité d'informatique et au Comité des méthodes et normes, pour approbation officielle.

Finalement, il est recommandé que le financement des logiciels généraux (développement, évaluation des progiciels commerciaux et soutien compris) se fasse au niveau du Bureau. Il sera déterminé par la demande des utilisateurs, mais non exclusivement celle d'un seul domaine d'activités. Auparavant, le gros des travaux effectués sur les logiciels généraux se faisait pour répondre aux demandes du recensement de la population. D'autres domaines d'activités, comme les enquêtes auprès des entreprises et l'analyse des données, ont également des besoins pressants de logiciels généraux. Il sera probablement plus facile d'économiser sur les ressources de développement si ces besoins étaient traités ensemble et non séparément.

Résumé de la discussion

Le statisticien en chef adjoint du Secteur de l'informatique et de la méthodologie, **M. J. Ryten,** *a donné un aperçu général des présentations sur l'automatisation du recensement.*

Introduction

L'automatisation peut s'appliquer à divers aspects du cycle d'exploitation du recensement. Les opérations de traitement qui comprennent la préparation du recensement, la saisie des données, le codage, le contrôle et l'imputation, les totalisations, la production de publications et de rapports d'information de gestion semblent le mieux convenir à l'automatisation. L'automatisation des opérations de traitement devrait accélérer la diffusion des données et réduire le personnel, mais risque d'avoir une incidence sur les coûts.

Application de l'automatisation

L'application de l'automatisation aux opérations de traitement du recensement devrait être fondée sur la détermination de la technologie la mieux adaptée à chacune des opérations qui composent la phase du traitement, par ex., l'opération saisie des données par rapport à l'opération contrôle et imputation. Le critère déterminant dans le choix d'une technologie tiendra compte de facteurs reliés à la fiabilité et à la polyvalence du système, et des coûts connexes.

Rôle des recensements passés et futurs

La recherche et le développement d'applications automatisées devraient également être basés et sur les résultats passés de l'application technologique et sur les conditions dans lesquelles se dérouleront les recensements futurs. Une certaine prudence serait de mise dans l'évaluation des résultats passés en raison du caractère exceptionnel et de la rareté des recensements. À chaque recensement, tout nouvel environnement technologique est une innovation.

Enfin, plusieurs des inquiétudes que suscitent les recensements futurs risquent d'augmenter, notamment les craintes du public à l'égard de la confidentialité et de la protection de la vie privée et les préoccupations du Bureau au sujet des restrictions budgétaires touchant les coûts de la main-d'oeuvre et du matériel ainsi que la polyvalence et les formes des technologies disponibles.

Résumé de la période de questions

Cette section contient un résumé des principales questions soulevées au cours de la présentation sur l'automatisation du recensement.

Questions touchant l'implantation de l'automatisation

D'abord, tout système automatisé a tendance à normaliser les données en supprimant les valeurs aberrantes. Cette normalisation peut compromettre la qualité des données de certaines variables.

Un système automatisé de codage est particulièrement utile dans le cas de variables telles que la profession ou l'industrie.

Il est fondé sur un dictionnaire ou une bibliothèque qui décrit les éléments de la variable. La Suède et le Canada (pour l'enquête sur la population active) ont développé de pareils systèmes automatisés de codage. Les deux systèmes diffèrent essentiellement par le taux de codage appliqué et le processus de mise à jour du dictionnaire ou de la bibliothèque.

Aux États-Unis, un système automatisé avait facilité la création d'un système efficace d'information de la gestion (SIG). Ce système permettrait d'améliorer considérablement le contrôle de gestion exercé sur les opérations de traitement.

L'utilisation de systèmes logiciels à la fois nouveaux et anciens peut causer des problèmes de fiabilité et d'incompatibilité. La réutilisation d'un système élaboré antérieurement est cependant encourageante à cause des facteurs coût et des ressources disponibles.

L'automatisation des opérations de traitement a soulevé deux questions touchant la participation des ressources humaines. La première question porte sur la nécessité croissante d'avoir un effectif spécialisé et expérimenté. La maintenance des logiciels existants et l'élaboration de nouveaux logiciels requièrent du personnel plus spécialisé que les opérations manuelles. Le statu quo en ce qui concerne l'allocation de ressources spécialisées limiterait la recherche et le développement en logiciel. La seconde question concerne la nécessité de réduire le nombre important d'employés temporaires sans expérience actuellement affectés aux opérations de traitement. La diminution de cet effectif contribuerait à réduire les coûts.

SÉANCE: COUVERTURE ET QUALITÉ DES DONNÉES

Président: Gordon Brackstone
Directeur général
Direction de la méthodologie

Jeudi, le 10 octobre 1985

CONDITIONS DE LA MESURE DE LA COUVERTURE ET DE L'AJUSTEMENT SUBSÉQUENT DANS LE CAS DES ÉTATS-UNIS

HOWARD HOGAN

STATISTICAL RESEARCH DIVISION
U.S. BUREAU OF THE CENSUS

Introduction

Depuis le premier recensement des États-Unis, en 1790, le repérage et le dénombrement précis de chaque personne ont toujours posé des problèmes. Les améliorations apportées aux techniques statistiques permettent aujourd'hui de déceler certaines erreurs de couverture, notamment un sous-dénombrement différentiel au niveau des groupes d'âge, du sexe, du groupe ethnique et de la région géographique. Devant l'évidence d'un tel sous-dénombrement différentiel et de sa persistance future, il a été proposé que l'on corrige les résultats des recensements au moyen de données statistiques appropriées. Pour bien comprendre la nécessité d'un tel ajustement, il convient d'aborder la question sous divers angles. En tout premier lieu, on doit se demander ce qu'est un recensement et ce que doivent en être les résultats. Ensuite, on posera des questions sur des aspects plus opérationnels: Comment pourrait-on améliorer les estimations du sous-dénombrement? Y a-t-il une meilleure façon d'établir les estimations de la couverture au niveau local et peut-on espérer obtenir des estimations qui soient d'une qualité suffisante? Doit-on rechercher la plus grande précision possible du dénombrement sur le terrain ou le meilleur ajustement possible? Sur quelles normes doit-on se fonder pour opter pour ou contre l'ajustement? Quelle pourrait être l'incidence du recours à l'ajustement sur la coopération du public et sur l'attitude des utilisateurs? Le présent document expose ces divers points et fait état des résultats de la recherche effectuée en la matière par le U.S. Census Bureau.

Si l'on considère le sous-dénombrement net, le recensement des États-Unis est en progrès constant. Une série d'estimations révèle que le taux de sous-dénombrement était supérieur à 4% pour 1950, d'un peu plus de 3% pour 1960, légèrement inférieur à 3% pour 1970 et d'environ 1% pour 1980. Autrement dit, au cours des quatre dernières décennies, le taux de sous-dénombrement est tombé de plus de 4% à 1%. Parallèlement à cette amélioration régulière du taux national, on note toutefois une persistance du sous-dénombrement différentiel. À chaque recensement depuis la Deuxième Guerre mondiale, le sous-dénombrement des Noirs américains a été de quelque 5% supérieur à la moyenne nationale. Pour ces quatre recensements, le sous-dénombrement des Noirs de sexe masculin a été d'au moins 7 points de pourcentage supérieur au sous-dénombrement à l'échelle nationale. Le recours envisagé à un ajustement des résultats du recensement n'est pas tant motivé par la préoccupation d'obtenir une bonne couverture globale du recensement que par le souci de corriger le sous-dénombrement différentiel persistant.

Il existe d'autres écarts au niveau de la couverture, bien qu'ils ne soient pas aussi bien étayés. Les villes qui constituent le coeur des grandes régions métropolitaines semblent enregistrer un sous-dénombrement plus important, dont l'importance décroît avec la taille de la ville et de la région métropolitaine. Les régions rurales connaissent également un sous-dénombrement considérable. Les taux de sous-dénombrement de certains groupes ethniques tels que les personnes d'origine hispanique, les Amérindiens ou les Asiatiques semblent plus élevés que la moyenne nationale, sans toutefois être aussi importants que pour les Noirs. L'expérience montre que le sous-dénombrement est plus élevé chez les personnes pauvres, les personnes seules et les chômeurs. Le sous-dénombrement qui touche les locataires semble plus élevé que celui qui caractérise les propriétaires occupants. On note un sous-dénombrement plus fort des personnes de moins de cinq ans et des personnes d'âge moyen, c'est-à-dire de vingt à quarante-cinq ans. Enfin, le taux de sous-dénombrement des hommes est régulièrement plus élevé que celui des femmes.

Le fait de savoir qu'il y a un certain sous-dénombrement et d'en connaître les tendances crée une incertitude; c'est comme de savoir qu'une montre avance ou recule légèrement. Cela ne dit pas ce qu'il faut faire pour qu'elle donne l'heure exacte. Auparavant, nos utilisateurs considéraient les résultats du recensement comme le reflet fidèle de la réalité ou, du moins, comme les meilleures données disponibles. À présent, ils doivent comme nous se demander ce qu'est un recensement et comment on pourrait l'améliorer. Ce problème est illustré dans la Constitution américaine qui prévoit d'une part que les sièges à la Chambre des Représentants doivent être alloués en fonction du chiffre global de population

mais aussi, d'autre part, qu'il faut procéder à un dénombrement réel. Dans l'esprit des auteurs, ces deux principes ne se contredisaient nullement car il n'existait pas alors d'autre possibilité que le dénombrement réel. Aujourd'hui, on peut avoir recours à l'estimation statistique et c'est pourquoi, pour la première fois, on est véritablement placé devant le problème.

Barbara Bailar a décrit les deux modèles de dénombrement, le modèle participatif et le modèle statistique. Dans le modèle participatif, on demande, pour obtenir des résultats précis, la participation des responsables des journaux ou revues, de la télévision, de la radio et celle des porte-parole des divers milieux. Le modèle participatif met l'accent sur l'importance intrinsèque et absolue de la participation, le recensement étant en quelque sorte assimilé à une grande cérémonie nationale. De toute évidence, la majorité des recensés qui remplissent leur questionnaire de recensement et aussi la majorité des utilisateurs ont ce modèle à l'esprit. Il est possible qu'à long terme le bien de la république soit le mieux assuré par une forte participation au recensement.

Néanmoins, un manque de précision des résultats du recensement nuit à tous, et non pas uniquement aux non-répondants. Le choix du modèle statistique repose sur la conviction que le bien de la république conduit à la nécessité d'avoir des résultats le plus exacts possible, même si les estimations produites sont basées sur des hypothèses arbitraires et non vérifiables. Au lieu d'employer les ressources uniquement à gagner la participation de toute la population, on utilise, dans le modèle statistique, une partie de ces mêmes ressources pour ajuster adéquatement les chiffres du recensement et produire des estimations valables.

Si l'on peut parler de deux modèles de recensement, on peut également considérer qu'il existe deux théories ou deux approches en ce qui concerne les résultats du recensement. Une théorie veut qu'on s'en remette au recenseur qui doit déclarer les données telles qu'elles lui sont fournies. L'autre veut que l'on utilise des chiffres de population estimatifs, qu'ils soient directement tirés des résultats du recensement ou non. Cette dernière approche élargit le champ de la question de l'ajustement bien au-delà du simple ajustement des chiffres du recensement ou des données de recensement et même des caractéristiques des personnes oubliées. Cela implique l'ajustement des caractéristiques déclarées en vue de corriger les inexactitudes connues. Par exemple, si les dossiers immobiliers révèlent que tous les immeubles d'un îlot ont été construits avant 1940, à quoi cela servirait-il de publier des données erronnées fournies par les résidents actuels concernant l'âge de ces immeubles? Les meilleures estimations sont basées sur les meilleures données.

Jusqu'à la Deuxième Guerre mondiale, presque tous les recensements ont été basés sur le modèle participatif et les résultats obtenus étaient indubitablement des données de recensement. Depuis la Deuxième Guerre mondiale, il y a eu une tendance, du moins dans les pays industrialisés, à utiliser des estimations plutôt que les chiffres réels. Imputation selon la méthode dite du "hot-deck", substitution ou attribution de données, correction des données se rapportant à la population mobile, correction des données en rapport avec l'occupation ou l'inoccupation, toutes ces formules sont tirées du modèle statistique. Cependant, l'évolution dans cette direction a été timide et le mouvement est tout juste amorcé. L'attrait pour l'ajustement statistique dans les cas de sous-dénombrement semble résulter non d'une volonté de procéder à des ajustements marginaux mais d'une évolution rapide vers l'adoption d'une nouvelle approche et d'une nouvelle méthode de dénombrement. Nos connaissances actuelles nous permettent-elles véritablement de corriger les données du recensement? Les progrès accomplis permettent-ils de produire des estimations statistiques de qualité supérieure à celles qui peuvent être obtenues dans le cadre d'une participation maximale de la population?

Dénombrement plus efficace ou estimations plus adéquates

Supposons un instant que le seul objectif du concepteur d'un recensement soit d'obtenir les meilleurs chiffres estimatifs de population possible pour les régions de petite taille. Quelles quantités de ressources doivent alors être consacrées au dénombrement sur le terrain et à l'évaluation de la couverture respectivement? Tout d'abord, que faut-il entendre par "ressources"?

Bien sûr, une ressource importante est l'argent. Les fonds pour le programme d'évaluation de la couverture et d'ajustement des données peuvent être prélevés du budget alloué au dénombrement sur le terrain. De fait, dans la mesure où l'on est assuré d'un minimum de fonds, les ressources financières constituent rarement une contrainte véritable en ce qui a trait à l'évaluation de la couverture et le montant requis n'est généralement guère élevé par rapport au coût de base du dénombrement. En 1980, les résultats n'auraient pas été plus précis quand bien même le

budget prévu pour le programme postcensitaire aurait été doublé. La difficulté que pose le maintien d'une qualité élevée des interviews et du couplage des données dans le cadre de l'évaluation de la couverture limite sérieusement la taille de l'échantillon. Cette contrainte se fait sentir bien avant que les coûts totaux deviennent importants par rapport au coût de base du dénombrement.

Quelles sont les autres contraintes? La première en importance est sans nul doute le temps. La question des délais revêt deux aspects différents. D'abord, il faut du temps pour effectuer le dénombrement sur le terrain. Plus on dispose de temps, plus grand est le nombre de personnes que l'on est en mesure de dénombrer, toutes choses égales d'ailleurs. Mais le processus d'évaluation de la couverture prend lui aussi du temps: interviews, couplage des données, dépistage et estimation. Comme il importe que les résultats du recensement soient prêts, disons, neuf mois après le jour du recensement, le temps dont on dispose est fixe. Le concepteur doit alors décider de la meilleure façon de répartir le temps alloué en vue de produire des résultats de la meilleure qualité possible. Un autre aspect qui doit être pris en considération est l'intervalle entre la date du dénombrement et la période de référence. Plus le jour du recensement est proche de la date de référence, meilleurs seront les résultats. De même, plus tôt l'évaluation de la couverture est effectuée, meilleures sont les réponses obtenues et plus la qualité des données est grande.

L'autre ressource indispensable est le personnel; il faut en particulier du personnel de bureau et du personnel sur le terrain qui aient reçu la formation voulue. Si le concepteur décidait de mener de front le dénombrement et une étude de mesure de la couverture, il faudrait recruter un double effectif tant en ce qui concerne le personnel de bureau que le personnel sur le terrain. Aux États-Unis, nous avons du mal à recruter un simple effectif temporaire pour le recensement.

Pour résoudre ce problème, les États-Unis et l'Australie ont adopté deux solutions diamétralement opposées. En 1980, on a tenté aux États-Unis de régler le problème du sous-dénombrement en concentrant tous les efforts sur le dénombrement initial. Des procédures spéciales ont été mises sur pied à cet effet. Les bureaux sur le terrain sont restés ouverts cinq mois et plus après le jour du recensement. Cela a conduit à une baisse notoire du sous-dénombrement moyen, qui a touché également la population noire. En Australie, tout au contraire, le dénombrement sur le terrain a été effectué en deux semaines! L'opération a été immédiatement suivie d'une enquête postcensitaire qui a abouti à un ajustement des données de recensement. Bien entendu, le dénombrement de certaines régions des États-Unis pose des difficultés qui n'existent pas en Australie et les deux cas ne sauraient se prêter à une comparaison valable. Néanmoins, il est bon de se demander si l'une des deux méthodes permet d'obtenir des estimations de qualité supérieure et, également, s'il existe une solution intermédiaire qui serait optimale. Il nous faut en outre étudier dans quelle mesure les essais et la recherche peuvent nous permettre de trouver une réponse empirique à ces questions.

Pour 1990, aux États-Unis, nous allons opter pour une position intermédiaire. La procédure adoptée consiste en un dénombrement intégral de la population concomitant à une évaluation de la couverture. Nous souhaitons obtenir les meilleures données possible à l'occasion de la collecte sur le terrain et établir les meilleures estimations possible avant la date fixée par la loi, soit le 31 décembre 1990. Nous comparerons ensuite les deux séries de données afin de déterminer dans quelle mesure les données ajustées répondent aux normes que nous avons fixées. Le choix de cette politique tient à ce que nous n'avons aucun moyen de savoir ou de prédire de façon valable si les données ajustées seront conformes aux normes établies ou même si elles seront produites dans le délai imparti. Un premier essai de cette double approche doit être effectué en 1986. Il s'agit d'un recensement d'essai à grande échelle concomitant à une évaluation de la couverture. Il est également prévu de produire des totalisations et des publications à cette occasion. Bien entendu, l'essai ne vise pas à déterminer le degré de précision de l'ajustement, étant donné que cela ne peut être évalué à partir d'un échantillon restreint. Le recensement d'essai nous permettra toutefois de vérifier la faisabilité des opérations d'ajustement.

Meilleure façon d'obtenir des estimations

Il ne s'agit pas ici de passer en revue toutes les méthodes possibles, avec leurs points forts et leurs lacunes. Depuis 1980, nous avons soigneusement examiné chacune des principales méthodes: enquête postcensitaire, contre-vérification des dossiers, appariement avec les registres administratifs, analyse démographique. Cet exposé reprendra donc seulement quelques points essentiels.

En premier lieu, le choix de la méthode d'évaluation la plus appropriée est lié au type de pays, d'administration publique, de recensement et de sous-dénombrement. L'importance de la population, le degré de mobilité de celle-ci de

même que la mesure dans laquelle les gens connaissent bien leurs voisins influent sur le résultat de la contre-vérification des dossiers aussi bien que sur celui de l'enquête postcensitaire. Dans un pays comme les États-Unis, où l'administration publique est décentralisée, les registres administratifs ont tendance à être dispersés. Les législations d'États ou locales en regard de la confidentialité obstruent ou retardent l'accès aux registres administratifs. Cette situation rend difficile la vérification des registres administratifs puisqu'aux États-Unis les registres des naissances sont conservés par les États plutôt que par l'administration fédérale. Conséquemment, la contre-vérification des dossiers à l'aide d'un échantillon des naissances devient plus ardue.

Il ne faut pas négliger de prendre en considération le type de recensement ni le type de sous-dénombrement lorsqu'on veut déterminer la meilleure procédure à adopter. La principale erreur de couverture est-elle due à des omissions, ou l'erreur nette est-elle à la fois le fait d'omissions et d'un certain sous-dénombrement? À quel intervalle de la date de référence la mesure de la couverture peut-elle être effectuée? Les omissions sont-elles attribuables principalement à une mauvaise application des procédures de dénombrement ou sont-elles aussi attribuables à la volonté de certaines personnes de se soustraire au recensement? Cette dernière attitude ne peut que donner des résultats biaisés lorsque la méthode employée implique une communication directe avec la personne en question. Il convient de souligner que la qualité de la contre-vérification des dossiers dépend de la qualité du dénombrement précédent et de celle de la mesure de la couverture précédente.

Le choix de la meilleure méthode de mesure dépend du problème en cause. L'expérience accumulée aux États-Unis a révélé que l'analyse démographique est souvent la façon la plus sûre de mesurer le sous-dénombrement à l'échelle nationale. Comme on pouvait s'y attendre, l'analyse démographique est également la façon la plus sûre de mesurer le sous-dénombrement dans le cas de groupes démographiques. Tout ce que nous savons concernant le sous-dénombrement de la population noire adulte de sexe masculin ou des enfants noirs est dérivé presque exclusivement de l'analyse démographique.

La contre-vérification des dossiers offre des avantages appréciables lorsqu'il s'agit de dériver des estimations du sous-dénombrement pour de grandes régions géographiques. Grandes s'entend, bien entendu, du point de vue du taux de migration interne. Dans le cas des régions de petite taille, comme les îlots de recensement, il est plus difficile de tirer un échantillon représentatif de la population actuelle que de la population d'il y a cinq ou dix ans. L'enquête postcensitaire est probablement la méthode la plus appropriée pour comprendre ce qui s'est passé dans un îlot moyen lors du dénombrement. Elle permet de repérer toutes les personnes qui résidaient dans l'îlot le jour du recensement, de déterminer qui a été dénombré et qui a été oublié, ou encore qui a été dénombré par erreur. Elle est également la meilleure façon de mesurer l'erreur globale de dénombrement, à un coût relativement bas. Néanmoins, l'enquête postcensitaire à elle seule ne permet pas d'évaluer un certain type d'erreur systématique, comme l'omission des personnes seules noires de sexe masculin.

Les registres administratifs restent probablement la meilleure solution en ce qui concerne certains groupes. Ainsi, puisque la mesure de la couverture vise à déterminer la qualité du dénombrement et le type de sous-dénombrement, le recours aux registres administratifs est une solution à envisager. Pour tout groupe particulier, conducteurs, anciens combattants, contribuables ou autres, il existe une série de registres administratifs à partir desquels on peut évaluer le dénombrement. Le problème, peut-être insurmontable, que pose l'usage des registres administratifs tient à la difficulté d'intégrer les diverses parties consultées de façon à constituer un échantillon aléatoire représentatif de la population globale, notamment de l'ensemble de la population d'une région de petite taille. Erikson et Kadane pensent avoir surmonté la difficulté grâce à l'approche des "listes composites" qu'ils ont mise au point. Toutefois, il ne nous semble pas que cela soit vraiment la solution.

Dans le cas des États-Unis, les problèmes occasionnés par le sous-dénombrement sont tels qu'une seule méthode ne saurait suffire à les régler. Il faut de toute évidence obtenir des estimations directes pour un échantillon de données régionales qui serviront dans l'application du modèle. Par ailleurs, il existe également un sérieux problème de sous-dénombrement systématique, comme nous l'avons souligné plus haut. Nous cherchons à améliorer l'enquête postcensitaire car, pour 1990 du moins, il nous semble qu'une variante d'une enquête-ménage est l'application la plus pratique à grande échelle dans le cas des régions de très petite taille. Nous étudions également la possibilité de combiner l'enquête postcensitaire avec d'autres méthodes telles que l'analyse démographique ou la vérification à l'aide des registres administratifs,

en vue d'améliorer les estimations relatives aux groupes plus difficiles à dénombrer.

Voyons maintenant les résultats d'une partie de nos recherches relatives aux méthodes d'évaluation de la couverture. Nous ne passerons pas en revue ici toutes les méthodes étudiées mais nous nous limiterons à quelques-unes des plus intéressantes.

La Forward Trace Study est une expérience qui vise à vérifier la faisabilité de la contre-vérification des dossiers dans un pays comme les États-Unis. L'étude a été entreprise en 1980 et le travail sur le terrain vient d'être terminé ce mois-ci. En 1960, nous avons établi que le fait d'attendre jusqu'après le recensement suivant pour commencer le dépistage conduisait à un taux de non-dépistage trop élevé. À cause de l'étendue considérable du pays, de la grande mobilité de la population et de la longueur de la période intercensitaire (dix années), le dépistage devenait trop difficile. Nous avons donc entrepris une expérience de dépistage immédiat. Au lieu d'attendre jusqu'à la fin de la période de recensement, nous procédons immédiatement au dépistage.

La Forward Trace Study en cours est une expérience à trois phases. Pour la phase du contrôle, nous avons adopté les méthodes utilisées au Canada, soit attendre jusqu'à la fin pour rassembler les données concernant le dépistage. Dans une autre phase, nous faisons les interviews dès le début pour recueillir des renseignements sur les personnes à contacter, parents ou amis, et pour obtenir le numéro d'assurance sociale de façon à pouvoir retrouver la personne dans les registres administratifs. Dans la troisième phase, les intéressés sont interviewés périodiquement durant le dépistage afin que leur trace ne soit pas perdue. Pour les deux groupes de dépistage actif, nous consultons les registres administratifs et postaux afin de mettre à jour les adresses.

L'étude porte sur les diverses parties de l'échantillon de la contre-vérification des dossiers qui est tiré à partir des personnes dénombrées, des naissances, des immigrants et des personnes oubliées lors du dernier recensement. À l'intérieur de l'échantillon, un coefficient de pondération élevé est attribué aux groupes pour lesquels l'enquête postcensitaire donne habituellement de moins bons résultats. En l'absence d'un recensement quinquennal, aucun appariement n'est possible; toutefois, nous pouvons évaluer notre capacité de conserver les mêmes personnes dans l'échantillon sur une longue période. Nous n'avons pas encore les résultats finals, mais il semble que le dépistage immédiat actif produise des résultats significatifs et devrait être envisagé dans un pays où l'on a recours à la contre-vérification des dossiers. Nous n'avons toujours pas pu déterminer si le dépistage est efficace.

Nous procédons à une expérience en vue d'établir si l'usage des registres administratifs peut permettre d'améliorer la couverture d'une enquête postcensitaire. Depuis 1960, d'énormes progrès ont été faits au niveau des méthodes de dénombrement grâce à l'adoption d'une variété de techniques permettant d'améliorer la couverture. Cependant, notre effort de recherche en ce qui a trait à l'enquête postcensitaire est limité au dénombrement conventionnel. Nous cherchons à utiliser les registres administratifs pour améliorer la couverture de l'**enquête postcensitaire**. Nous espérons ainsi englober une partie de la population touchée par l'omission systématique.

Jusqu'à maintenant nous n'avons parlé que de l'**enquête postcensitaire**. Il est également possible d'effectuer une enquête **précensitaire**. C'est une façon d'éviter quelques-unes des contraintes de temps. Le choix de cette méthode s'appuie également sur la nécessité d'habituer le public et le personnel sur le terrain à cette technique. Une expérience d'envergure modeste portant sur l'enquête précensitaire de 1986 est en cours de planification.

Enfin, la partie la plus intéressante de notre travail concerne l'appariement automatisé et l'appariement manuel assisté par ordinateur. L'automatisation du recensement de 1990 a ouvert des possibilités d'appariement par ordinateur qui n'avaient encore jamais existé. On imagine facilement les gains de temps et d'argent que cela permet. Cette méthode présente aussi l'avantage d'améliorer considérablement la précision de l'appariement et de réduire le biais consécutif à l'appariement. Notre équipe de recherche sur l'appariement a mis au point un programme d'appariement automatisé d'une puissance surprenante, en appliquant la théorie élaborée par Fellegi et Sunter. Nous mettrons ce programme à l'essai pour la première fois avec les résultats de l'enquête postcensitaire expérimentale de 1985. Pour 1986, notre objectif est d'intégrer ce système à un système d'appariement manuel en direct avec l'ordinateur afin de pouvoir résoudre rapidement les cas d'appariement probable et les autres cas problèmes.

Estimations régionales

L'importance d'un recensement tient, en grande partie, à ce qu'il permet d'avoir des données pour

de très petites régions. Nombre d'utilisations des données de recensement requièrent des estimations portant sur des régions de très petite taille: révision des districts électoraux, allocation de fonds, prévision des effectifs scolaires et planification du réseau d'autoroutes, pour n'en nommer que quelques-unes. Pour ce qui est des États ou des provinces, des estimations moins onéreuses sont statistiquement, sinon constitutionnellement, tout aussi acceptables.

C'est ici qu'on est confronté à un dilemme. Les méthodes que nous utilisons pour évaluer le sous-dénombrement sont d'autant plus efficaces que la région visée est importante: un État, une grande agglomération ou l'ensemble du pays. Autrement dit, il nous est possible de savoir s'il y a eu sous-dénombrement dans le cas des grandes régions géographiques. Il nous est également possible de déterminer l'importance du sous-dénombrement dans le cas d'une strate de régions semblables. Néanmoins, les meilleures estimations de la population que nous ayons pour chaque îlot sont probablement les chiffres recueillis sur le terrain par le personnel du recensement.

En fait, ce paradoxe, pratique plutôt que statistique, est particulièrement marqué lorsqu'il s'agit de surdénombrement. Il est parfaitement possible, sur la base des faits et des modèles observés, de conclure à un surdénombrement pour un certain type de régions. Pourtant, les chiffres de population les plus précis pour chaque îlot de la région en question sont probablement ceux qui ont été recueillis sur le terrain.

La recherche actuelle dans ce domaine est orientée vers l'utilisation d'estimateurs synthétiques intégrés à un plan d'échantillonnage en vue d'en maximiser l'efficacité par rapport au problème de sous-dénombrement. Au lieu de nous concentrer sur la production d'estimations directes pour les États et les grandes agglomérations, nous envisageons de procéder à des regroupements d'États de façon à former des strates homogènes du point de vue des caractéristiques qui nous paraissent liées au sous-dénombrement. Ce peut être des caractéristiques de logement, la composition raciale ou ethnique, le revenu ou encore l'éducation. Nous nous basons sur les résultats du recensement de 1980 pour étudier une telle approche. Comme on ne dispose pas d'estimations du sous-dénombrement pour les régions de petite taille, dans un premier temps nous substituons des données de recensement, c'est-à-dire que nous faisons une imputation de toutes les données relatives à la personne, à titre de variable de remplacement. Nous avons opté pour la substitution de données car, non seulement les données en question ont un lien avec le sous-dénombrement, mais encore elles ont la même importance. Nous pouvons alors confronter le taux estimé par cette approche avec les opérations relatives aux résultats de recensement connus.

Normes d'ajustement

Ce choix suppose implicitement que l'on peut déterminer si les données ajustées sont meilleures que les données recueillies sur le terrain. Mais qu'entend-on au juste par "meilleures"? Pour répondre à cette question, il faut utiliser une fonction de perte permettant de comparer les deux séries de données. Supposons que l'on dispose d'une série de données qui correspond au meilleur dénombrement possible sur le terrain. Supposons également que l'on dispose par ailleurs d'une autre série de données qui correspond au meilleur ajustement possible. Supposons enfin, en étant peu réalistes, que les données véritables sont connues. Il est peu probable que l'une ou l'autre des deux séries sera plus conforme aux données véritables pour toutes les régions et tous les groupes sans exception. Par conséquent, on doit pouvoir se servir d'une fonction de perte pour pondérer les erreurs, même si ce n'est que de façon théorique.

L'erreur quadratique moyenne, l'erreur absolue relative et la borne inférieure de l'erreur sont des exemples de fonctions de perte. Idéalement, il faut baser le choix de la fonction sur les utilisations prévues des données. Autrement dit, la forme de la courbe de perte doit refléter la perte réelle pour la société occasionnée par les erreurs de recensement. La tâche est loin d'être aisée, même si l'on considère une seule utilisation à la fois. Vouloir en donner une représentation adéquate pour toutes les utilisations principales représente un défi de taille.

Bien entendu, on ne peut définir des normes en vue de l'ajustement à partir de l'hypothèse que les données réelles sont connues. Les décisions doivent être fondées sur des données observables. Il peut s'agir des erreurs de couverture mesurées, du niveau du sous-dénombrement différentiel ou des taux de non-réponse relevés à l'occasion de l'évaluation de la couverture. Ces données observables doivent correspondre à des erreurs existant dans les deux séries de données qui supposent des niveaux donnés de la fonction de perte. Une grande partie de la recherche consiste donc à trouver une façon de relier les fonctions de perte. Cela revient, dans un premier temps, à établir les profils d'erreur pour les estimations tirées de la mesure de la couverture et pour celles qui sont tirées du recensement. Cela suppose également l'élaboration de méthodes pour évaluer

les erreurs ou, du moins, leur ordre de grandeur. Enfin, cela implique que l'on trouve un moyen de mettre en évidence le lien entre les erreurs mesurées et les fonctions de perte, en ayant recours à l'analyse ou à des simulations. Le United States Census Bureau a entrepris une recherche poussée dans ce sens. Nous cherchons à obtenir l'aide de statisticiens répartis dans toute l'Amérique du Nord. La tâche est considérable mais tout progrès effectué constituera un pas vers une meilleure compréhension de la nature particulière de ce type de décision.

Il va de soi, comme nous l'avons rappelé plus haut, que le recensement n'est pas un exercice purement statistique. Quelle risque d'être l'incidence de l'ajustement sur la participation du public? L'expérience à ce chapitre est limitée; on dispose toutefois d'un certain nombre de faits utiles. En Australie, paraît-il, l'usage de l'ajustement n'a entamé en rien la collaboration de la population. Aux États-Unis, les gens ne semblent pas déplorer le fait que nous imputions les données relatives au nombre d'occupants des logements lorsque ce nombre n'a pu être obtenu. En 1970, les chiffres de population des États-Unis ont été grossis par imputation à la suite d'une étude statistique portant sur les personnes oubliées parce que les logements qu'elles habitaient avaient par erreur été déclarés inoccupés. Cela n'a nullement indisposé le public. À notre sens, l'important c'est que la population considère l'ajustement comme une méthode de recensement normale. Par contre, si l'ajustement était perçu comme devant remplacer le recensement, comme une façon de modifier celui-ci ou comme tout autre moyen pour en réduire l'importance, les réactions risqueraient alors d'être très fortes.

De même, les résultats du recensement doivent être présentés le plus simplement possible. Si, comme il se doit, on part de l'hypothèse que tout ajustement doit viser à améliorer la qualité des données, l'utilisateur en sera immanquablement satisfait. L'ignorance est parfois une garantie de sérénité, mais elle peut aussi conduire à des décisions coûteuses. Les utilisateurs aiment qu'il y ait cohérence dans le contenu des tableaux et dans les résultats. Pour avoir des tableaux suffisamment cohérents, les données doivent être ajustées jusqu'à un niveau très élémentaire. De fait, la façon la plus simple de procéder peut être d'imputer les personnes oubliées au niveau des microdonnées du recensement. Cela pose toutefois un problème. Nous savons évaluer le sous-dénombrement de la population (du moins en théorie) et imputer des données en conséquence. Nous savons également comment évaluer le sous-dénombrement des ménages complets, même si cela s'avère plus difficile, et imputer les données en conséquence. Par contre, nous ne savons comment faire les deux en même temps tout en préservant la cohérence des résultats. Par exemple, comment doit-on procéder dans le cas de personnes oubliées à l'intérieur d'un ménage?

Il ne fait aucun doute que ces problèmes peuvent être résolus et qu'ils le seront, quoique cela ne sera peut-être pas pour 1990. Faute d'une solution intégrale, on peut mettre au point des moyens de présenter les données d'une façon utile et significative. Cependant, l'ajustement constitue un changement auquel l'utilisateur doit s'adapter. Il incombe à l'organisme chargé du recensement de prévoir un programme en vue d'expliquer ce changement aux utilisateurs des données.

Conclusion

Le United States Bureau of the Census n'a pas encore défini de normes pour l'ajustement, et encore moins décidé si un ajustement peut permettre de respecter de telles normes. Mais, il s'est donné pour mission de mener à bien un programme énergique de recherche visant à fixer de telles normes et, dans la mesure du possible, de respecter ces dernières. Ce programme nous a déjà valu la contribution efficace de Statistique Canada. Nous espérons grandement que les résultats de cette recherche de même que nos observations à ce propos contribueront à faciliter la planification du recensement du Canada de 1991. Nous avons nous-mêmes déjà eu l'occasion de profiter de votre expérience.

VERS UNE ÉVALUATION PLUS APPROPRIÉE DE LA QUALITÉ DES DONNÉES

RICHARD BURGESS

DIVISION DES MÉTHODES DE RECENSEMENT ET D'ENQUÊTES-MÉNAGES
STATISTIQUE CANADA

Introduction

L'évaluation de la qualité des données et le rôle de celle-ci dans le recensement ont toujours été et continuent d'être un sujet de discussion et de travail de premier ordre. Environ 5% du coût des recensements de 1981 et de 1986 correspondent aux opérations liées à l'évaluation de la qualité des données. Cependant, malgré cela il n'est pas certain que les coûts impliqués ni les activités spécifiques continueront d'être justifiés. Nous aborderons donc la question de la pertinence de l'évaluation de la qualité des données dans le cadre du recensement canadien.

Nous verrons d'abord le contexte particulier dans lequel les activités de mesure et d'évaluation de la qualité des données doivent être planifiées et mises sur pied. Puis, nous donnerons une description des activités courantes sous ce rapport et nous exposerons les attentes des utilisateurs vis-à-vis du programme d'évaluation de la qualité des données. Enfin, nous ferons un certain nombre de suggestions visant à améliorer la pertinence d'une telle évaluation.

Contexte du recensement

La mesure de la qualité des données du recensement est effectuée dans un contexte unique. Ces dernières années, une des principales caractéristiques de ce contexte a été celle du manque de ressources. Les ressources sont limitées non seulement du point de vue du budget ou des années-personnes utilisables mais encore de celui du personnel expérimenté et spécialisé disponible.

Deuxièmement, au cours des années le processus de recensement a été de plus en plus orienté vers la production. Les problèmes et les changements sont perçus avant tout et essentiellement comme une menace par rapport au respect du calendrier et (ou) du budget. La grande majorité du personnel chargé du recensement (il y a, par exemple, quelque 40,000 recenseurs) doit appliquer un ensemble de procédures dans un temps limité. Ce personnel est appelé à prendre des décisions sur le champ et, même s'il est bien intentionné, il arrive fréquemment qu'il ne soit pas en mesure d'évaluer avec justesse l'incidence des décisions ainsi prises sur la qualité du produit final.

Troisièmement, il ne faut jamais perdre de vue que les données recueillies à l'occasion du recensement doivent avoir une utilité. Dans certains cas, cet impératif va à l'encontre du besoin de données de haute qualité. La variable "type de construction", recueillie en 1981, est un bon exemple de cette contradiction. La répartition détaillée des codes a produit des données de piètre qualité. Par ailleurs, le regroupement des codes qui permet d'avoir des données de meilleure qualité donne des chiffres beaucoup moins utiles.

Enfin, que ce soit au moment de la planification, de la mise en oeuvre ou de l'évaluation, il faut toujours avoir à l'esprit que Statistique Canada est responsable de la qualité et de l'utilité des données de recensement.

Mesure de la qualité des données sous sa forme actuelle

Connaissant le contexte, nous allons maintenant analyser l'évaluation de la qualité des données sous sa forme actuelle.

En ce qui a trait aux problèmes techniques, un autre exposé de cette séance (donné par Howard Hogan) présente d'excellents exemples de quelques-uns des problèmes liés à la mesure de la qualité. Ces problèmes débordent le cadre de la simple mesure de l'erreur de couverture bien que, pour le recensement, plus de 75% des dépenses au titre de l'évaluation de la qualité s'appliquent à la couverture.

Il est reconnu internationalement que la mesure de la qualité des données doit être effectuée et que les résultats doivent en être diffusés. À Statistique Canada, cette exigence fait partie intégrante de la politique officielle. Pour ce qui est du recensement, l'analyse de la qualité des données est effectuée à la fois par des études directes et au moyen d'un processus de certification. La combinaison de ces deux méthodes vise à mieux faire comprendre ce que les données représentent réellement.

Toutefois, les activités touchant la qualité des données sont parfois considérées comme une cause possible de non-respect du calendrier. Il y a toujours un risque que la diffusion des données soit retardée si un problème a été repéré et qu'un nota de mise en garde ne soit pas prêt dans les délais voulus. Il y a aussi le fait que les questionnaires de recensement sont très demandés. Au moment même où les données sont saisies, les personnes chargées de la contre-vérification des dossiers, de la vérification des logements inoccupés et d'autres études d'évaluation essaient d'avoir accès aux données du recensement. Dans un tel contexte, la mesure de la qualité des données n'est pas toujours perçue comme faisant partie intégrante du recensement proprement dit. Elle l'est, par contre, lorsque quelque chose va mal ou que quelqu'un a critiqué les données.

En ce qui touche l'utilisation faite des mesures de la qualité des données, il est certain qu'il y a quelques problèmes réels. Ainsi, l'intégration formelle des mesures de la qualité des données au processus de prise de décision peut souvent compliquer singulièrement la tâche. Pour cette raison, on a tendance à laisser de côté ces mesures, même lorsqu'il s'agit d'une mesure aussi élémentaire que l'erreur d'échantillonnage. Il est beaucoup plus simple de prendre les données telles quelles et d'effectuer une analyse directe. Ceux qui produisent des mesures de la qualité des données s'accordent pour dire que, bien souvent, les utilisateurs ne sont pas intéressés à connaître la qualité des données. Ces derniers sont surtout préoccupés d'avoir des données. Même s'ils étaient intéressés par les mesures de la qualité, ils ne disposent généralement pas des moyens techniques nécessaires pour utiliser les mesures que nous produisons.

Compte tenu de ce qui précède, il faut reconnaître que nous ne procurons pas aux utilisateurs l'aide dont ils ont besoin pour bien utiliser les mesures de la qualité des données. Ainsi, bien que nous produisions une quantité considérable de renseignements sur l'erreur de couverture, nous ne faisons aucun ajustement global en fonction de cette erreur de couverture. Que nous attendons-nous que les utilisateurs fassent des mesures que nous leur fournissons? Nous nous contentons d'encourager, en termes généraux, l'utilisation des mesures de la qualité des données. Dans certains cas, comme pour le taux de réponse, c'est la seule chose que l'on puisse faire. Il se peut aussi que ce soit précisément ce que l'utilisateur recherche, en ce sens que bon nombre des utilisateurs veulent seulement savoir si les données sont utilisables sans plus. Cependant, dans la plupart des cas, nous laissons l'utilisateur décider de lui-même ce qu'il doit faire des mesures de la qualité des données.

Enfin, il convient de souligner que les études sur la qualité des données représentent le plus souvent un travail considérable, notamment en termes du nombre de personnes qu'il faut y affecter. Ces personnes doivent concilier les impératifs de la demande de mesures de la qualité des données et la nécessité de procéder à des analyses approfondies. La production de telles mesures ne saurait être assimilée à un simple processus de production, car les résultats obtenus doivent être d'une précision rigoureuse. Ces derniers doivent être étayés par une documentation très claire et sûre.

Besoins des utilisateurs en matière de renseignements sur la qualité des données

Les utilisateurs des résultats de l'évaluation de la qualité des données du recensement sont nombreux et ont chacun leurs besoins propres. Les observations qui suivent sont plutôt spéculatives car il est en pratique difficile de savoir exactement quels renseignements veulent les utilisateurs.

Les utilisateurs de données, c'est-à-dire les personnes qui utilisent les données produites à partir des résultats du recensement, souhaitent avant tout avoir un avis clair et facile à comprendre sur la qualité des données qui leur sont fournies ou une mesure simple et d'emploi facile de la qualité de ces mêmes données. Ils ont besoin notamment d'être avisés de la qualité des données régionales et d'être conseillés sur la façon d'utiliser les mesures et renseignements correspondants.

Les gestionnaires et les planificateurs du recensement constituent un deuxième groupe d'utilisateurs. La haute direction et les planificateurs peuvent avoir à se pencher sur des questions comme la nécessité d'effectuer un contrôle postal et dans quelles conditions ou la nécessité d'avoir recours à l'autodénombrement ou à un dénombrement anticipé dans certains cas bien spécifiques. Les réponses à ces questions dépendent en bonne partie de la qualité visée. Le plus souvent, il est impossible de fournir une réponse claire du fait qu'on ne dispose pas de renseignements adéquats sur l'incidence probable de la décision envisagée sur la qualité des données.

Les planificateurs sont particulièrement préoccupés d'assurer la bonne qualité des données du recensement. Une fois le processus de recensement en marche, rien ne peut l'arrêter. Par

conséquent, le niveau de qualité est déterminé par les données fournies par les recensés, ce qui veut dire qu'il dépend essentiellement de la qualité du questionnaire et des procédures de collecte des données. La qualité repose également sur la façon dont le personnel temporaire affecté au recensement (c'est-à-dire un nombre important de personnes) traite les données recueillies. En d'autres mots, il ne suffit pas d'élaborer des procédures adéquates, il faut également vérifier que celles-ci sont correctement appliquées.

Un troisième groupe d'utilisateurs est constitué par les membres du personnel de divers services de Statistique Canada. Les personnes chargées d'effectuer les analyses et d'élaborer les spécifications du contenu pour le prochain recensement doivent avoir une bonne connaissance des données et, notamment, de leur qualité. Le personnel opérationnel doit également être au courant de l'incidence de son travail sur la qualité des données, cela ne pouvant qu'accroître sa motivation. Par exemple, le personnel des bureaux régionaux se sert des estimations de l'erreur de couverture pour évaluer le travail effectué dans le cadre d'un recensement particulier. Dans la mesure où le personnel chargé du dépouillement au bureau régional ou au bureau central est au fait de l'incidence sur la qualité des données des décisions qu'il est amené à prendre, il est sans doute porté à réfléchir ou à s'informer davantage avant de prendre une décision finale.

Enfin, la mesure de la qualité des données profite à l'ensemble de Statistique Canada dans ce sens qu'elle confirme l'intégrité de celui-ci. Il est bon, lorsque des gens de l'extérieur critiquent les données, de pouvoir dire que nous avons examiné la qualité même si nous devons admettre que celle-ci est limitée.

Amélioration de la pertinence de l'évaluation de la qualité

Du fait que nous consacrons beaucoup de ressources à la mesure de la qualité des données, nous devons nous demander s'il serait possible d'accroître la pertinence d'une telle évaluation.

Tout d'abord, nous avons besoin de plus de renseignements sur les besoins des utilisateurs, sur ce que ces derniers sont en mesure de comprendre et sur les types d'analyses auxquelles les données sont soumises. Il est difficile de fournir des mesures appropriées aux utilisateurs si nous ne savons pas ce qu'ils veulent en faire. À l'occasion du recensement de 1991, cet aspect pourrait être inclus dans le processus de consultation en rapport avec le programme de production de données.

Nous pouvons également améliorer les modes de présentation des données sur la qualité. Le seul fait d'adopter une présentation simplifiée, par exemple, en utilisant des graphiques, permettrait de retenir plus sûrement l'attention des utilisateurs.

En ce qui a trait aux moyens que nous utilisons actuellement pour diffuser les données sur la qualité, nous pourrions envisager d'autres possibilités pour remplacer les publications volumineuses et complexes actuelles. Nous pourrions avoir à la place une sorte de guide sommaire d'une vingtaine de pages au maximum qui résumerait brièvement les méthodes utilisées en ce qui a trait à la qualité des données du recensement et donnerait ensuite, pour chaque variable, un aperçu de notre point de vue sur les données: tout à fait utilisables, de valeur discutable, etc. Dans certains cas, nous pourrions renvoyer les utilisateurs à d'autres sources ou les aviser qu'ils doivent se procurer d'autres renseignements concernant telle variable. Cette approche permettrait d'atteindre davantage de gens et aurait plus d'effet que celle que nous avons adoptée jusqu'à présent.

Nous aurions également avantage à fournir quelque indication de la variation de qualité d'une région géographique à une autre. Jusqu'à présent, nous avons dans la plupart des cas présenté une série de résultats assortie de la mention: "le taux réel peut varier d'une région à une autre". Cela traduit bien la réalité mais ce n'est guère utile. Une telle solution pourrait même passer pour une dérobade.

Nous devrions aussi tenter d'intégrer les mesures de la qualité des données à nos propres activités d'analyse. D'aucuns objecteront que nous ne faisons pas suffisamment d'analyses ou que cela en ferait augmenter le coût. Nous pourrions cependant joindre une sorte de résumé ou de condensé pour montrer aux utilisateurs comment ils peuvent tirer parti de ces renseignements dans leurs analyses.

Enfin, nous pourrions mettre davantage en évidence le fait que la mesure de la qualité des données influe sur la forme que prendra le prochain recensement. Cet aspect a été très difficile à justifier au cours des derniers recensements par suite de la limitation des ressources. Il est relativement difficile de convaincre le chef de projet du recensement de 1986 de prendre à son compte une activité qui n'aura de résultats qu'en 1991. Néanmoins, dans le passé cet aspect a été un des meilleurs arguments en faveur de l'évaluation de la qualité des données. Les mesures de la qualité ont conduit, par exemple,

à l'utilisation de l'autodénombrement, de l'échantillonnage de 1 sur 3 et de 1 sur 5 et, dans une grande mesure, à l'élaboration des systèmes CANEDIT et SPIDER. Il est très important que nous nous consacrions davantage à ce type d'activité car il semble que nous avons atteint un palier en ce qui a trait à la qualité des données.

En pratique, il y aurait plusieurs manières pour nous d'atteindre cet objectif. Nous pourrions, par exemple, aller sur le terrain et déterminer la cause réelle des erreurs. Cela est particulièrement important pour ce qui est de la couverture, car dans ce domaine les données dont nous disposons traduisent les symptômes et non la cause réelle. Non seulement ces renseignements seraient utiles pour améliorer le prochain recensement mais encore ils faciliteraient l'ajustement. Nous pourrions en outre élargir la portée de nos études actuelles sur la qualité des données, notamment de celles qui portent sur la couverture. Nous pourrions en profiter pour recueillir des renseignements sur d'autres sujets, par exemple, en demandant aux gens s'ils comprennent bien la question sur la langue maternelle ou sur l'origine ethnique et s'ils estiment pouvoir bien répondre à partir des catégories de réponse proposées dans le questionnaire.

En résumé, il convient de reconsidérer l'importance et l'orientation du programme relatif à la qualité des données. Ce programme constitue certainement un investissement considérable, mais l'adoption de quelques modifications mineures et d'une nouvelle perspective permettrait de produire des mesures plus utiles pour les utilisateurs de l'extérieur comme pour les planificateurs du Bureau.

AJUSTEMENT DANS LE CAS DES ERREURS AUTRES QUE LES ERREURS DE COUVERTURE

CHRIS HILL

DIVISION DES MÉTHODES DE RECENSEMENT ET D'ENQUÊTES-MÉNAGES
STATISTIQUE CANADA

Introduction

Les procédures adoptées dans le cadre du recensement canadien pour l'ajustement des données manquantes ou incohérentes sont parmi les plus évoluées et les plus complètes au monde. Nous verrons, dans le présent exposé, la raison d'être de ces procédures, nous en expliquerons les avantages et les inconvénients et nous exposerons les modifications qui mériteraient d'être apportées. Nous proposerons trois autres méthodes qui pourraient être adoptées pour le recensement de 1991. Nous ne traiterons pas en détail les méthodes ni les systèmes d'imputation utilisés pour le recensement puisque ceux-ci sont abondamment décrits dans d'autres documents.

Nécessité d'ajuster les données

Un ajustement est nécessaire dans le cas de données manquantes ou de données incohérentes. Une telle démarche fait partie intégrante du processus d'enquête. Il est nécessaire de passer des données brutes à des données en quelque sorte finales qui sont considérées comme des estimations officielles ou du moins suffisamment proches de la réalité. Diverses méthodes peuvent être utilisées pour ce faire: contrôle et imputation, c'est-à-dire ajustement des microdonnées dans les enregistrements, application d'un nouveau coefficient de pondération, voire suppression des enregistrements erronés. Il importe de reconnaître que la suppression des enregistrements erronés représente en fait une forme d'ajustement des données.

Idéalement, on vise à obtenir des estimations le plus justes possible. En pratique, on obtient une base de données exempte des inconvénients dus aux données manquantes et aux données incohérentes. Y a-t-il des choses que nous pouvons améliorer? En réalité, bon nombre d'erreurs ne peuvent être détectées au niveau des microdonnées. Ces erreurs n'apparaissent qu'à un stade ultérieur. Certaines erreurs peuvent biaiser les résultats, d'autres pas; on ignore malheureusement, au moment de produire les données du recensement, quel est ce biais. Lorsque les résultats de l'évaluation sont disponibles, il est trop tard pour procéder à des ajustements. Du fait que nous procédons à l'ajustement de certaines données mais non d'autres, il convient de le faire avec modération. Nous sommes justifiés de nous demander pourquoi nous consacrons tant de ressources et d'énergie à une opération qui ne résout que partiellement le problème.

À titre d'exemple, prenons le cas des données sur la langue maternelle. En 1971 et en 1981, nous avons publié des données sur la langue maternelle ajustées en fonction des cas de non-réponse. En 1976, nous avons maintenu la catégorie "Non déclaré". Au moment de la publication des données de 1976, les journalistes de la presse anglophone aussi bien que ceux de la presse francophone en ont conclu que, pour la première fois dans l'histoire du Canada, le pourcentage de personnes dont la langue maternelle est le français est tombé en dessous de 80%. Or, si l'on effectue un ajustement même grossier en fonction des réponses de la catégorie "Non déclaré" publiées en 1976, on observe au contraire pour le Québec un accroissement général du pourcentage de personnes dont la langue maternelle est le français.

La question n'est pas de savoir si les données doivent ou non être ajustées. Des ajustements sont indispensables dans le cas des données manquantes et probablement aussi pour les données incohérentes. La véritable question est de savoir si Statistique Canada doit se charger de cette tâche ou si celle-ci peut être confiée à d'autres.

Arguments contre l'ajustement fait par Statistique Canada

Voyons tout d'abord pourquoi Statistique Canada ne devrait pas avoir la responsabilité de l'ajustement des données.

Une des raisons est d'ordre financier. L'ajustement des données est une opération coûteuse. En 1981, elle a représenté environ 10% du coût global du recensement.

Une autre raison qui a probablement encore plus de poids est la nécessité de procéder à l'ajustement dans les délais les plus courts possible. En 1981, il a fallu ajouter quatre mois au processus de

recensement pour l'ajustement des données intégrales. Dans le cas des données de l'échantillon de 20%, le délai a été d'environ un an. Dans ce contexte, les quelques semaines nécessaires pour la collecte des données sont insignifiantes.

Le troisième point à considérer est le caractère aléatoire du processus. Aux recensements de 1971 et de 1976, la publication des données a été sérieusement retardée à cause de problèmes survenus au moment du contrôle et de l'imputation. En 1981, cela a failli être le cas à plusieurs reprises.

Le quatrième argument est le fait que nous avons été accusés de manipuler les données. Dans divers articles et à l'occasion de diverses conférences auxquelles j'ai assisté, l'opinion a été exprimée que nous n'étions pas intègres et que nous publions des données arrangées de façon à paraître meilleures. Les auteurs de telles critiques se rangent en deux catégories. Certains prétendent que nous manipulons les données d'une façon que les utilisateurs ne peuvent déceler. Les autres admettent que les utilisateurs comprennent ce que nous faisons mais condamnent le type d'ajustement que nous faisons. Un exemple particulièrement frappant est le cas des données sur la langue pour lesquelles nous avons procédé à un certain nombre d'ajustements qui ont été très critiqués.

Arguments en faveur de l'ajustement fait par Statistique Canada

On peut, de la même façon, avancer de solides raisons pour que Statistique Canada continue d'effectuer l'ajustement des données. Le premier argument est d'ordre qualitatif. Le Bureau peut faire le meilleur ajustement possible du fait que son personnel a seul accès aux microdonnées et peut donc tenir compte de tous les renseignements qui y sont contenus au moment de procéder à l'ajustement. Parallèlement à cela, le Bureau a accumulé une expérience considérable et s'est donné une compétence indiscutable en la matière.

Le deuxième argument repose sur le besoin de maintenir la compatibilité des données et comporte trois aspects. Il s'agit, tout d'abord, de préserver la compatibilité à l'intérieur des tableaux et entre les divers tableaux pour un même recensement. En l'absence d'ajustements appropriés, les divers tableaux peuvent présenter certaines incohérences qui risquent d'occasionner des ennuis aux utilisateurs. Dans un deuxième temps, il convient de préserver la compatibilité des données chronologiques. Cela commande que le Bureau soit chargé de l'ajustement des données mais aussi qu'il s'acquitte de cette tâche de manière uniforme. L'exemple des données sur la langue maternelle que nous avons utilisé plus haut montre clairement les dangers d'un manque de cohérence. Enfin, les utilisations des données du recensement étant régies par la loi et les règlements administratifs, il importe que les données publiées soient cohérentes du point de vue de chacun des intéressés en cause.

La troisième raison pour laquelle l'ajustement aurait avantage à être fait par Statistique Canada est qu'il s'agit d'un service aux utilisateurs. Nombre d'utilisateurs des données ne tiennent pas à avoir à faire face aux problèmes et aux complications occasionnés par des données incomplètes. Il est sans doute plus pratique que nous nous chargions de faire l'ajustement une seule fois pour tous les utilisateurs plutôt que de laisser ces derniers le faire chacun individuellement. Par ailleurs, à l'intérieur même du Bureau, le fait que les personnes chargées de la diffusion des données soient aux prises avec des problèmes dus à un manque de cohérence des données qui doivent être produites constitue un handicap majeur.

Enfin, un dernier argument en faveur de l'ajustement fait par le Bureau est la nécessité pour celui-ci de préserver sa crédibilité. La présence d'anomalies dans les séries produites risquerait, pensent certains, d'entamer la confiance que les données fournies inspirent aux clients. Personnellement, je ne pense pas qu'un tel argument soit valable.

Trois options possibles pour 1991

Compte tenu des diverses raisons pour ou contre l'ajustement des données par Statistique Canada, trois options sont envisageables pour 1991. La première consiste à poursuivre notre façon de faire actuelle mais en accordant plus d'attention à ce que l'on considère être la "gestion des risques". La deuxième option repose sur la publication préliminaire de données non vérifiées. Et, enfin, la troisième suppose que l'ajustement effectué est limité à quelques variables seulement; autrement dit, il excluerait certaines des variables qui ont fait l'objet d'un ajustement jusqu'à présent.

Pour ce qui est de notre façon de faire actuelle, il est intéressant d'en rappeler rapidement les principales particularités. Depuis le recensement de 1971, Statistique Canada a eu largement recours au contrôle et à l'imputation automatisés. De nombreux problèmes s'étant posés en 1971, on a mis sur pied le système CANEDIT afin de les éviter en 1976. Ce système comporte deux caractéristiques essentielles. Il s'appui sur une

méthodologie rigoureuse et il permet de contrôler étroitement les processus d'élaboration et de mise en oeuvre. Toutefois, nous avons établi en 1981 que le système CANEDIT imposait certaines contraintes arbitraires qui entamaient la flexibilité des ajustements possibles dans certains domaines. Nous avons également établi que certaines variables, principalement celles qui sont assorties d'une longue liste de codes ou les variables arithmétiques, ne pouvaient pas être traitées par CANEDIT. Pour ces diverses raisons, nous avons donc élaboré un nouveau système, le SPIDER. Les quatre points précités, à savoir la rigueur, le contrôle, la flexibilité et le choix de variables pouvant être traitées, sont des caractéristiques indispensables à un système de contrôle et d'imputation. La question est de savoir si un même système peut les offrir toutes simultanément.

En ce qui a trait plus spécifiquement à la gestion des risques, voici quatre recommandations. Premièrement, il conviendrait de réutiliser CANEDIT en 1991. Deuxièmement, il faudrait rendre le système SPIDER plus rigoureux. Troisièmement, il serait souhaitable d'éliminer certaines des complexités propres aux stratégies de contrôle et d'imputation. Enfin, il serait nécessaire d'analyser plus en détail notre façon de contrôler l'ensemble du processus. Bien que nous ayons accompli beaucoup de progrès en 1981 en ce qui a trait au contrôle de l'ensemble du système d'imputation, il faut encore harmoniser davantage cette composante.

Voyons maintenant la solution plus controversée de la publication préliminaire de certaines données qui, de toute évidence, romperait la pratique établie en ce qui concerne le recensement. Il y a déjà eu un ou deux précédents, soit la publication anticipée des totalisations spéciales pour le Yukon et les Territoires du Nord-Ouest; mais cela s'est fait à titre officieux. Toutefois, d'autres pays publient des données non vérifiées. Israël, par exemple, publie d'abord des données non vérifiées et ultérieurement les données vérifiées. L'Australie également pratique une forme de contrôle et d'imputation beaucoup moins stricte que ce que nous faisons et publie des données qui, selon nos critères, ne sont pas des données vérifiées.

Il est intéressant de souligner, par ailleurs, que bien qu'il soit de pratique courante à Statistique Canada de publier des chiffres provisoires suivis de chiffres ajustés pour ce qui est des variables économiques, cela ne s'est encore jamais fait dans le cas des variables sociales. Ces données sont publiées beaucoup plus tard et sous leur forme finale. Le principe qui sous-tend la publication de données provisoires mérite d'être analysé plus en détail dans le contexte du recensement. Douze mois est une bien longue période pour garder les statistiques dans la base de données, particulièrement lorsqu'un certain nombre d'utilisateurs savent pertinemment qu'elles y sont.

Bien entendu, la préoccupation qui surgit alors concerne le coût d'une telle pratique. Il importerait d'analyser les coûts additionnels liés à la publication de données provisoires ainsi que les problèmes logistiques que cela entraînerait. Or, dans l'état actuel des choses, nous produisons déjà un volume important de totalisations de contrôle en vue de déterminer si le processus d'ajustement est satisfaisant. Bon nombre de ces totalisations de contrôle pourraient être publiées. Il est suggéré que ces données soient diffusées, non pas dans une publication soignée ou coûteuse mais sous une forme relativement économique, à titre de données provisoires, avec un avis spécifiant qu'il s'agit de totalisations compilées rapidement à partir de données pouvant contenir des erreurs.

Une telle solution soulève le problème très réel de la crédibilité du Bureau. Dans le passé, l'objection a déjà été soulevée que le fait de publier des données qui comportent des erreurs risque de nuire à notre crédibilité. Or, à mon sens, l'argument inverse peut tout aussi bien être avancé. En effet, si les utilisateurs voient à la fois les données non vérifiées et les données vérifiées, ils pourront constater que le traitement que nous faisons subir à ces dernières n'a qu'une faible incidence sur leur teneur et que les ajustements que nous effectuons visent essentiellement à les rendre plus utiles et plus maniables et non à les modifier de façon majeure.

La troisième solution, qui consiste à réduire le nombre d'ajustements, est liée à la deuxième solution d'une certaine manière. Cependant, au lieu de publier des données provisoires qui seront vérifiées ultérieurement, nous pourrions envisager d'éliminer toute vérification subséquente ou de nous contenter d'un ajustement simplifié, peu coûteux et facile à faire, comme de pondérer les données en fonction des cas de non-réponse. Une telle approche offre des avantages évidents du point de vue des délais et des coûts.

Pour ce qui est de tenir compte de l'opinion des utilisateurs, il convient de séparer ces derniers en deux groupes. Un premier groupe correspond aux utilisateurs qui attendent seulement des chiffres sans plus de détail, généralement pour des régions de petite taille, et ne veulent pas avoir à se préoccuper d'explications complexes sur la façon d'utiliser les données. L'autre groupe correspond

à un petit nombre d'universitaires qui doivent faire des analyses très complexes des données qui leur sont fournies. Les ajustements détaillés que nous effectuons actuellement ne satisfont aucun des deux groupes. Pour les utilisateurs du premier groupe, les explications complexes que nous offrons sont inutiles; un ajustement plus simple correspondrait tout autant à leurs besoins. Les utilisateurs plus exigeants sont, eux aussi, insatisfaits car ils n'approuvent pas nécessairement nos méthodes et peuvent préférer avoir des données non vérifiées pour les analyser à leur guise. Dans l'état actuel des choses, en ce qui a trait aux caractéristiques culturelles du moins, notre pratique actuelle sert mal nos utilisateurs.

Résumé

En résumé, trois recommandations peuvent être faites en vue du recensement de 1991. Tout d'abord, il importe de concentrer les efforts sur la gestion des risques, notamment en évitant d'intégrer dans le processus d'ajustement des éléments qui ne sont pas véritablement nécessaires. Deuxièmement, il conviendrait d'envisager de produire des données provisoires au moins pour quelques figures-clés. Enfin, il faudrait voir sérieusement à limiter la portée des ajustements pour certaines variables qui font actuellement l'objet d'une telle opération.

REGISTRES DES ADRESSES: AVANTAGES ET INCONVÉNIENTS

DAVID C. WHITFORD

DECENNIAL PLANNING DIVISION
U.S. BUREAU OF THE CENSUS

Introduction

Pour le recensement décennal de la population et du logement de 1970, le U.S. Bureau of the Census a adopté, pour la première fois, la méthode de dénombrement d'envoi et de retour du questionnaire par la poste, dans la plupart des grandes régions urbaines. Dans ces régions, une liste des logements avait été établie et un questionnaire de recensement expédié à chaque adresse figurant sur la liste. Ces adresses ont ensuite été consignées dans des registres des adresses qui ont servi de moyen de contrôle du dénombrement. Chaque registre des adresses (RA) contient les renseignements relatifs aux logements d'une région géographique donnée (environ 300 logements par registre). Durant le recensement, les RA ont servi à corriger les adresses, à en ajouter de nouvelles ou à en éliminer ainsi qu'à consigner le nom du responsable du ménage, la description de l'emplacement, la date de réception du questionnaire, les chiffres de population et la désignation du contenu de l'échantillon.

Le U.S. Census Bureau a utilisé trois sources générales pour compiler les listes d'envoi qui ont permis de constituer les registres des adresses: des listes ont été dressées à l'occasion des visites sur le terrain, d'autres listes ont été achetées de vendeurs privés et les listes établies lors des recensements antérieurs ont été mises à jour.

La deuxième étape du processus d'élaboration des registres des adresses consiste à mettre à jour ces listes une fois qu'elles sont prêtes. Les opérations de mise à jour, conçues de façon à assurer l'intégralité (ou la couverture) de l'inventaire des adresses domiciliaires, sont de deux types. D'une part, les recenseurs vérifient sur le terrain toutes les adresses répertoriées et, d'autre part, les facteurs comparent les adresses figurant sur les listes du Census Bureau à celles de leur itinéraire respectif.

Enfin, les registres des adresses sont utilisés pour contrôler la livraison des questionnaires de recensement. Ces derniers sont envoyés par la poste ou remis à un recenseur qui les livre aux adresses répertoriées dans le registre tout en mettant celui-ci à jour par la même occasion. Le recensement peut également être mené sans établir de registres des adresses avant le dénombrement. En 1970 et en 1980, pour les régions des États-Unis où l'on n'a pas utilisé les services postaux, le recensement a été fait à partir de listes d'adresses constituées lors du dénombrement sur le terrain.

Le présent document compare les diverses méthodes d'établissement et de mise à jour des registres des adresses ainsi que les moyens d'en assurer l'intégralité (ou la couverture), les modes de livraison et les méthodes de dénombrement possibles; il donne également un aperçu des divers usages des registres des adresses.

Pour chacun de ces éléments, la comparaison portera sur les points suivants: déroulement des opérations et traitement des RA, amélioration de la couverture, coûts, avantages et inconvénients.

Nota:

- Un graphique de cheminement des opérations est joint au texte pour chaque méthode d'établissement et de mise à jour des listes d'adresses et pour chaque méthode de livraison des questionnaires afin de montrer de quelle manière la méthode en question peut être intégrée au processus de recensement. Ces graphiques sont repris en annexe pour illustrer les diverses façons possibles d'élaborer les registres des adresses.

- Les données sur la couverture et sur les coûts qui sont présentées ici proviennent de sources diverses et correspondent à des circonstances différentes. Il n'entre pas dans le cadre du présent document de donner les précisions propres à ces données, telles que les niveaux de variance et les particularités de l'échantillonnage. Il n'est pas recommandé d'établir des généralisations applicables au recensement national à partir des données sur les coûts et la couverture qui sont présentées ici.

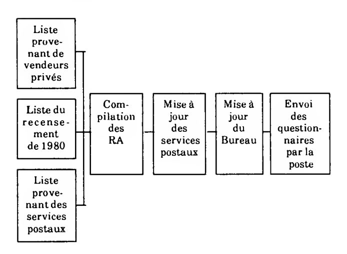
Acquisition des listes préliminaires

Le U.S. Bureau of the Census a considéré quatre sources différentes pour l'acquisition des registres des adresses:

- achat de listes de vendeurs privés,
- achat de listes des services postaux des États-Unis,

- mise à jour des registres du recensement précédent,
- développement de listes à l'occasion des opérations sur le terrain.

En ce qui a trait aux préparatifs d'envoi des questionnaires par la poste, pour les trois premières méthodes le graphique de cheminement des opérations est le suivant:



La compilation des adresses pour constituer un registre nécessite le regroupement des logements par région géographique (géocodage), le traitement des cas ne pouvant être géocodés et l'impression des registres. Cela fait, les registres des adresses sont prêts à être mis à jour.

Il convient de souligner qu'entre l'acquisition des listes et l'opération d'envoi et de retour du questionnaire par la poste deux mises à jour sont effectuées: l'une par les services postaux et l'autre par le Census Bureau. Un des objectifs essentiels poursuivis dans le cadre de l'établissement des listes d'adresses en vue d'un recensement postal est de faire coïncider l'adresse utilisée par le Census Bureau et l'adresse postale. En d'autres mots, toute adresse postale doit pouvoir être repérée sur le terrain par le recenseur chargé d'interviewer le ménage correspondant.

Dans notre test de compilation des listes d'adresses (ALCT, Address List Compilation Test), nous avons comparé récemment la couverture et les coûts des listes achetées de vendeurs privés, des listes fournies par les services postaux et des listes du recensement de 1980 (mises à jour)[1]. Pour une seule région urbaine, la couverture et les coûts de la liste initiale se présentent comme suit:

Liste	Couverture: adresses valides	Coût: par adresse
Vendeur privé	48,640	$0.05
Services postaux	49,988(2.8%)	0.97
Recensement de 1980	50,752(4.3%)	0.15

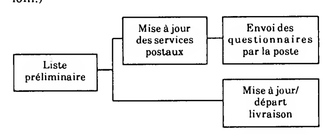
Nota:

- Le test de compilation des listes d'adresses a été effectué en 1984. La liste tirée du recensement de 1980 manquerait grandement d'actualité si elle était utilisée pour le recensement de 1990.

- Comme nous l'avons déjà souligné, les données sur la couverture qui sont présentées ici correspondent aux listes originales. Elles ne font pas état des améliorations qui pourraient résulter de la mise à jour. (Voir la section suivante à ce sujet.)

La quatrième façon d'acquérir les listes d'adresses consiste en un repérage sur le terrain par le personnel du Census Bureau. Pour ce faire, on délimite la région géographique et un recenseur parcourt le secteur en inscrivant à la main dans les registres des adresses les coordonnées de chaque logement. (Ces renseignements sont ensuite entrés par clavier en vue du traitement automatisé.) En 1980, un listage préliminaire a ainsi été effectué dans les régions pour lesquelles on n'avait pu obtenir de listes de vendeurs privés.

Dans les régions où le dénombrement est effectué à partir de listes préliminaires, la livraison des questionnaires peut être faite de deux façons: par la poste ou par livraison à domicile, assurée par le Census Bureau une fois que les listes ont été mises à jour. (Cette dernière méthode sera exposée plus loin.)



Une mesure de la couverture des listes obtenues de vendeurs privés et des listes préliminaires est donnée par le taux des ajouts résultant de la mise à jour ultérieure par les services postaux. En 1980, les taux obtenus pour les deux types de listes étaient sensiblement les mêmes [2].

[1] Franklin, D., Dinwiddie, J. et Lueck, M., 11 juillet 1985. "Results and Analysis of the Urban Address List Compilation Test", note de service adressée par Charles D. Jones à Susan M. Miskura.

[2] Whitford, D. et Thomas, K., 30 juin 1983. "Post Office Effectiveness", résultats de l'évaluation préliminaire du recensement de 1980, note de service nº 52.

Le listage préliminaire est une opération coûteuse. En 1980, le coût a été de $1.39 par logement[3].

Il convient de souligner qu'en 1980 les listes préliminaires n'étaient pas soumises à la vérification postale préalable effectuée dans les régions couvertes par les listes obtenues de vendeurs privés. Nous verrons cette vérification en détail un peu plus loin.

Avantages et inconvénients des méthodes d'acquisition des listes préliminaires:

Le coût semble être l'élément le plus pertinent en ce qui a trait à l'établissement de listes aux États-Unis.

- D'après les résultats du test de compilation des listes d'adresses, la compilation est plus onéreuse dans le cas des listes obtenues des services postaux que pour les autres listes.

- L'utilisation des registres des adresses établis à l'occasion d'un recensement antérieur est une approche prometteuse dans la mesure où les coûts liés à la mise à jour des listes d'un recensement à un autre ne deviennent pas prohibitifs.

- La nécessité du repérage sur le terrain (listage préliminaire) est apparente lorsqu'on ne dispose d'aucune autre liste ou lorsque les listes accessibles ne sont pas adéquates.

Mise à jour des listes préliminaires

Les opérations de mise à jour des listes préliminaires sont présentées ci-après. La mise à jour vise un double objectif:

- faciliter la livraison des questionnaires en cas de recensement utilisant la méthode de retour par la poste,

- vérifier la couverture de la liste originale.

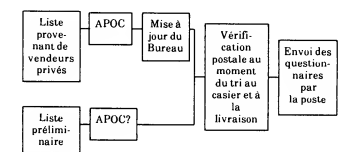
Comme nous l'avons déjà souligné, le problème que pose l'utilisation de listes préliminaires vient de ce que les adresses utilisées par le Census Bureau et par les services postaux ne correspondent pas toujours. En outre, d'autres problèmes particuliers à ce type de listes, comme le fait que des logements inoccupés ont pu devenir occupés, que de nouveaux logements ont pu être construits, que des logements ont pu être démolis ou encore que le numéro d'appartement dans un immeuble à logements multiples a pu être mal défini, imposent une mise à jour des listes préliminaires avant leur compilation initiale.

Le U.S. Bureau of the Census a adopté trois processus pour la mise à jour des registres des adresses. Ces processus sont présentés dans les deux sections suivantes qui portent sur la mise à jour par les services postaux et la mise à jour sur le terrain.

Mise à jour par les services postaux

La vérification préalable par les services postaux (APOC, Advanced Post Office Check) a été appliquée au recensement de 1980 à titre de vérification initiale des listes achetées de vendeurs privés.

Cette mise à jour a été effectuée à l'été 1979. Elle peut être insérée dans le cheminement des opérations du recensement de la façon suivante:



On remarquera que, dans le cas des régions pour lesquelles on a établi des listes préliminaires, il y a un point d'interrogation à côté de la vérification préalable par les services postaux (APOC). Cela signifie qu'au recensement de 1980 il n'y a pas eu de vérification des listes préliminaires par les services postaux. Ce type de vérification initiale des adresses des listes préliminaires fait actuellement l'objet d'un essai en vue de son éventuelle application au recensement de 1990.

Pour ce qui est de la vérification préalable par les services postaux, une carte d'adresse a été remplie pour chacune des adresses figurant dans les listes obtenues de vendeurs privés et envoyée au bureau de poste. Les adresses domiciliaires auxquelles la poste livre du courrier et qui ne figuraient pas dans les listes du Census Bureau ont été enregistrées sur des "cartes

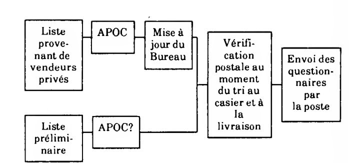
[3] Sans auteur, 3 décembre 1982, "1980 Decennial Census Costs".

bleues" par les facteurs. Ces adresses ont ensuite été géocodées (lorsque cela était possible), vérifiées afin d'éviter toute duplication et, lorsqu'elles étaient valides, elles ont été incluses dans les registres des adresses.

L'efficacité de la vérification préalable par les services postaux est reflétée par le taux de couverture et le montant des coûts engagés. Cette vérification a permis d'ajouter 5.5 % du nombre total des logements dénombrés dans les régions pour lesquelles les listes ont été obtenues de vendeurs privés. Quatre-vingt-treize pour cent de ces logements se sont avérés occupés[4]. Au départ, les coûts de la vérification préalable par les services postaux ont été fournis non par logement mais par logement ajouté. Chaque adresse ajoutée aux registres à la suite de cette vérification a coûté $3.49 au Census Bureau[5].

Juste avant le 1er avril, jour du recensement des États-Unis, lorsque les questionnaires ont été envoyés au bureau de poste, deux autres opérations de mise à jour ont été effectuées, la vérification au moment du tri au casier et la vérification à la livraison des questionnaires. En 1980, la vérification au moment du tri au casier a été entreprise quelque trois semaines avant le jour du recensement. Là encore, des "cartes bleues" ont été envoyées au Census Bureau pour tout logement qui n'avait pas eu de questionnaire. Les adresses correspondantes ont été ajoutées aux registres après vérification de leur validité. Puis, au moment de la vérification à la livraison des questionnaires, le facteur devait consigner tout logement qui n'avait pas eu d'enveloppe du recensement. On a ensuite vérifié si les adresses correspondantes figuraient dans les registres et, dans la négative, elles y ont été ajoutées lorsqu'elles étaient valides.

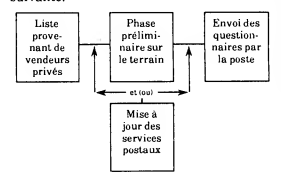
La vérification au moment du tri au casier et la vérification à la livraison des questionnaires peuvent être insérées dans le cheminement des opérations du recensement de la façon suivante:



La vérification au moment du tri au casier et la vérification à la livraison des questionnaires ont été effectuées dans les régions où l'on avait opté pour l'envoi et le retour du questionnaire par la poste. Cette double vérification a permis une amélioration, estimée à 3.4%, de la couverture des ménages dans ces régions. Quatre-vingt-neuf pour cent des logements ainsi ajoutés se sont avérés occupés[6]. Le coût de ces vérifications a été de $2.56 par logement ajouté[7].

Mise à jour par le Census Bureau

Du fait qu'il existait un problème de cohérence entre les adresses postales et celles qui sont utilisées par le Census Bureau, ce dernier a procédé à sa propre mise à jour de janvier à mars 1980. La période avait été choisie de façon à tomber entre les deux mises à jour effectuées par les services postaux. Cette mise à jour interne, également appelée phase préliminaire des opérations sur le terrain, s'insère dans le cheminement des opérations de recensement de la façon suivante:



Durant la phase préliminaire des opérations sur le terrain, un employé du

4 Whitford, D. et Thomas, K., 30 juin 1983, op. cit.

5 Thompson, J., 28 août 1984. "Preliminary Summary Results from the 1980 Census Coverage Improvement Program Evaluations", résultats de l'évaluation préliminaire du recensement de 1980, note de service n° 85.

6 Whitford, D. et Thomas, K., 30 juin 1983, op. cit.

7 Thompson, J., 28 août 1984, op. cit.

Census Bureau a utilisé le registre des adresses (vérifié par les services postaux) pour se rendre à chaque logement répertorié pour la région donnée. Il devait consigner les changements observés et l'adresse des nouveaux logements, rayer l'adresse des logements qui n'existaient plus, préciser certains détails relatifs à l'emplacement et inscrire le nombre d'appartements dans les immeubles à logements multiples. Pour 1990, le Census Bureau a mis à l'essai un type de dénombrement préliminaire au cours duquel les coordonnées de chaque logement de tout immeuble à logements multiples sont consignées dans le registre des adresses et font l'objet d'une vérification particulière.

À la fin de la phase préliminaire des opérations sur le terrain, les changements relevés sont saisis en vue de la publication d'une nouvelle version du registre des adresses et les documents d'expédition sont produits en tenant compte des mises à jour.

Il est difficile d'obtenir des données en vue d'établir si la phase préliminaire des opérations sur le terrain a permis d'améliorer la couverture du recensement de 1980 étant donné qu'à la veille du jour du recensement plusieurs opérations étaient effectuées en même temps dans les bureaux de recensement. On estime toutefois à quelque 2.36 millions le nombre des logements ajoutés aux registres des adresses à cette occasion. Le coût par logement ajouté a été d'environ $5.00[8].

Avantages et inconvénients des diverses techniques de mise à jour:

La détermination de la date des mises à jour et la coordination de la parution des listes d'adresses révisées par le Census Bureau et par les services postaux semblent être des préoccupations majeures en ce qui a trait à l'utilisation de la mise à jour pour élaborer les registres des adresses.

- La vérification préalable par les services postaux, qui consiste à mettre à jour les listes obtenues de vendeurs privés dès leur acquisition, est une vérification initiale servant à préparer les adresses en vue de leur mise à jour finale par le Census Bureau et par les services postaux.

- La phase préliminaire des opérations sur le terrain permet de résoudre les incohérences entre les adresses postales et les coordonnées sur le terrain en vue des activités de suivi postcensitaire.

- Les vérifications au moment du tri au casier et à la livraison des questionnaires sont effectuées quand les services postaux ont reçu les questionnaires à livrer. C'est pourquoi il semble approprié d'avoir une vérification ultime à cette occasion également.

Méthodes de livraison

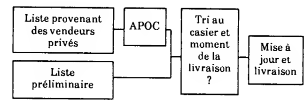
Les méthodes de livraison des questionnaires aux logements se classent en deux catégories: celles qui nécessitent la compilation de listes d'adresses préalablement au dénombrement (présentées sous forme de registres des adresses) et celles qui supposent l'élaboration des registres des adresses à l'occasion des visites sur le terrain effectuées par les recenseurs. Nous avons vu plus haut les méthodes d'envoi et de retour des questionnaires par la poste. Dans la présente section, nous traiterons de la méthode qui combine la mise à jour des adresses et la livraison des questionnaires et de celle qui combine le listage des adresses et la livraison des questionnaires. Nous présenterons également les résultats des analyses de la couverture et des coûts effectuées par le U.S. Bureau of the Census pour chacune des méthodes précitées.

Mise à jour des adresses et livraison des questionnaires

La méthode combinant la mise à jour des adresses et la livraison des questionnaires consiste à faire **réviser** par les employés du Census Bureau les registres des adresses (établis avant le recensement) au moment où les questionnaires sont **livrés** à tous les logements. Cette méthode a été utilisée, à titre expérimental, lors du recensement de 1980.

Cette méthode s'inscrit dans le cheminement des opérations de la façon suivante:

[8] Fan, M. et Thompson, J., 22 octobre 1984. "Evaluation of the 1980 Census Precanvass Coverage Improvement Operations", résultats de l'évaluation préliminaire du recensement de 1980, note de service n° 92.



En quelques mots, les différences entre cette méthode expérimentale et celle qui est employée dans toutes les autres régions où l'on a recours au dénombrement postal sont les suivantes:

- la phase préliminaire des opérations sur le terrain a été supprimée et la mise à jour des registres des adresses est faite à l'occasion de la livraison des questionnaires,

- des cartes de rappel sont produites et expédiées aux logements.

Le plan adopté pour l'essai effectué en 1980 comportait cinq paires de bureaux de districts de recensement formées chacune de deux bureaux ayant les mêmes caractéristiques. Dans un bureau de chaque paire, on a remplacé la phase préliminaire des opérations sur le terrain (lorsqu'il ne s'agissait pas d'une région faisant l'objet de listage préliminaire) et la livraison postale par la méthode combinant la mise à jour des adresses et la livraison des questionnaires. Bien que dans les bureaux combinant la mise à jour des adresses et la livraison des questionnaires les questionnaires aient été livrés aux logements par un recenseur (ce qui donnait l'occasion d'inciter les recensés à remplir leur questionnaire), les résultats n'ont révélé aucune différence sensible au niveau de la couverture des ménages[9]. La méthode combinant la mise à jour des adresses et la livraison des questionnaires est plus onéreuse que l'envoi par la poste des questionnaires.

Listage des adresses et livraison des questionnaires

Dans les bureaux de recensement qui appliquent la méthode "conventionnelle", c'est-à-dire où aucune liste d'adresses préliminaire n'est achetée ni compilée, les procédures adoptées pour 1970 et 1980 sont les suivantes:

- expédition des questionnaires aux ménages sous forme d'un envoi global (c'est-à-dire sans adresse sur le questionnaire);

- désignation d'un recenseur chargé:

  - de parcourir les régions qui lui ont été attribuées, le jour du recensement ou les jours suivants,

  - de dresser la liste des logements qui s'y trouvent,

  - de reprendre les questionnaires remplis, de laisser un questionnaire en blanc ou d'interviewer le ménage à l'occasion de sa visite;

- application des procédures de mise à jour des adresses après le recensement, ce qui implique une vérification par les services postaux des adresses consignées lors du dénombrement.

Le cheminement de ces opérations est le suivant:



À l'occasion d'un programme expérimental mis sur pied dans le cadre du recensement de 1970, soit le test relatif à l'extension de la livraison postale des questionnaires, on a comparé le taux de couverture obtenu avec cette méthode aux taux de la livraison postale des questionnaires. À cet effet, cinq bureaux de recensement qui avaient été désignés pour appliquer la méthode conventionnelle sont devenus des bureaux d'envoi et de retour des questionnaires par la poste. Dans le plan de l'échantillon expérimental, ces bureaux ont été mis en paires avec cinq autres bureaux appliquant la méthode conventionnelle.

---

9 Bailey, L. et Ferrari, P., 15 juin 1984. "1980 Census Update List/Leave (ULL) Household Roster Check – Preliminary Report", résultats de l'évaluation préliminaire du recensement de 1980, note de service nº 70.

Les résultats obtenus sont les suivants:

- On n'a relevé aucune différence significative dans la couverture entre les deux catégories de bureaux.

- Dans les régions pour lesquelles une liste d'adresses préliminaire est dressée, le nombre d'adresses supprimées par les recenseurs et le nombre d'adresses ajoutées par les services postaux étaient directement proportionnels à l'erreur de couverture.

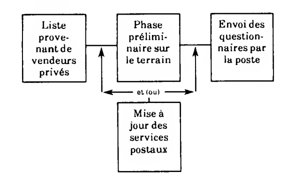
- En ce qui a trait aux données sur la population qui devaient être fournies par tous les recensés, les questionnaires provenant des régions faisant l'objet d'un recensement postal étaient plus complets.

- Au niveau des coûts, on n'a pas relevé de différence significative entre les deux méthodes[10].

Il convient de souligner que le test de 1970 a été appliqué dans toutes les régions rurales du pays. En 1980, le Census Bureau a repris l'approche combinant le listage des adresses et la livraison des questionnaires dans bon nombre de régions présentant les mêmes caractéristiques.

Livraison postale

Nous avons vu plus haut les procédures d'acquisition et de mise à jour des registres des adresses pour les régions des États-Unis où un listage préliminaire est effectué. L'acheminement des opérations pour les régions où l'on a recours à la livraison postale des questionnaires figure encore comme suit:



La livraison postale et la méthode combinant le listage des adresses et la livraison des questionnaires ont été comparées précédemment (test relatif à l'extension de la livraison postale des questionnaires de 1970). Des renseignements complémentaires sur la couverture doivent paraître dans un rapport sur le programme postcensitaire relatif au recensement de 1980. Ce rapport, qui donne une mesure de la couverture pour toutes les régions des États-Unis, est sur le point d'être publié. Toutefois, le fait que les caractéristiques géographiques et démographiques des régions où l'on applique la livraison postale des questionnaires et de celles où l'on applique la méthode combinant le listage des adresses et la livraison des questionnaires sont très différentes limite l'utilité de cette comparaison de couverture. Toute comparaison des coûts engagés pour chacun de ces deux types de régions présente également les mêmes limites. On dispose des données sur les coûts pour chaque type de région, pour le recensement de 1980. Dans tous les bureaux régionaux du Census Bureau sauf un, le coût des opérations sur le terrain par logement dans les régions où l'on a appliqué la méthode conventionnelle (listage des adresses et livraison des questionnaires combinés) est supérieur au coût enregistré pour les régions où l'on applique la livraison postale des questionnaires[11].

Avantages et inconvénients des diverses méthodes de livraison des questionnaires

Les avantages et les inconvénients de l'application de la méthode combinant le listage des adresses et la livraison des questionnaires et de la méthode d'envoi et de retour des questionnaires par la poste ne sont pas évidents. Une comparaison qui portait sur les régions rurales n'a pas révélé de différences significatives.

D'après l'expérience accumulée au Census Bureau des États-Unis, il apparaît que la méthode combinant le listage des adresses et la livraison des questionnaires est particulièrement appropriée aux régions rurales et aux populations éparses alors que dans les régions plus urbanisées il vaut

[10] U.S. Bureau of the Census, juin 1973. **Results and Analysis of the Experimental Mail Extension Test**, 1970 Evaluation and Research Program PHC(E).

[11] Imprimé d'ordinateur, 31 août 1983. "Table III. 1980 Census Field Operations Costs/Regional Office Costs by Type of District Office".

mieux effectuer le dénombrement à partir de registres des adresses établis au préalable.

Utilisation des registres des adresses

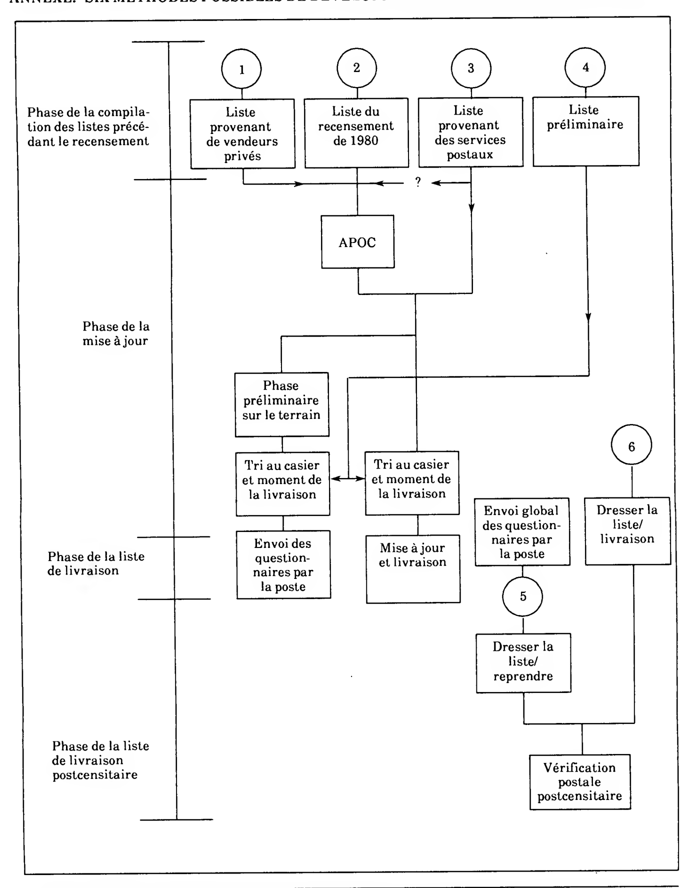
Les considérations qui précèdent suivent le cycle des registres des adresses en présentant les diverses méthodes de compilation, les stratégies d'amélioration et les rôles dans le cadre des diverses techniques de livraison des questionnaires.

Comme cela a déjà été souligné, les registres des adresses servent à consigner les données relatives aux logements et à contrôler certains aspects du dénombrement sur le terrain. Les données répertoriées comportent quelques-uns des éléments suivants sinon tous: code géographique, adresse (nom de la rue et numéro de la maison, numéro de l'appartement, ville, État, code régional), nom du responsable du ménage et description de l'emplacement. Ces données ainsi que les autres renseignements contenus dans les registres des adresses permettent de fournir ou de contrôler les éléments suivants:

- estimations de la charge de travail nécessaires pour planifier l'effectif, déterminer les besoins en services de soutien et en fournitures, établir les budgets, etc.;
- corrections, ajouts et suppressions d'adresses;
- données nécessaires à l'évaluation (par ex.; accroissement de la couverture des adresses résultant d'une procédure de mise à jour);
- état d'avancement de la réception ou de la non-réception des questionnaires nécessaire pour planifier le suivi;
- chiffres provisoires de population et des logements;
- contrôle qualitatif de la collecte des données auprès de l'échantillon.

En résumé, les registres des adresses sont un moyen de contrôler le dénombrement sur le terrain dans le cadre du recensement. L'utilisation de tels registres offre beaucoup de possibilités du point de vue de la gestion et du contrôle des activités.

ANNEXE: SIX MÉTHODES POSSIBLES DE DÉVELOPPEMENT DES REGISTRES D'ADRESSES



APPLICATIONS DU REGISTRE DES ADRESSES AU RECENSEMENT DU CANADA

DON ROYCE

DIVISION DES MÉTHODES DE RECENSEMENT ET D'ENQUÊTES-MÉNAGES
STATISTIQUE CANADA

Introduction

Une des activités-clés dans le cadre de la tenue du recensement de la population est le listage des logements de tout le Canada. En 1986, cette tâche doit être exécutée par le recenseur sur la base du secteur de dénombrement (SD) et les logements situés dans un SD donné doivent être répertoriés dans le Registre des visites (RV) correspondant.

Cette liste sert à divers usages au cours du recensement:

a) elle permet d'avoir une couverture complète et précise des SD étant donné que chaque logement donne lieu à une description unique;

b) elle sert de base d'échantillonnage au recenseur qui doit livrer un questionnaire complet à un ménage sur cinq;

c) elle sert à consigner l'état d'avancement pour chaque logement durant la collecte des données, c'est-à-dire qu'on y inscrit la date de livraison et la date de retour des questionnaires, l'avis de suivi lorsque cela est nécessaire, le nom de la personne à contacter;

d) elle sert de document de référence pour les contrôles (par exemple, chiffres de population et des logements) qui doivent être exécutés au cours du traitement subséquent des questionnaires;

e) elle permet de produire les chiffres provisoires de la population et des logements.

Comme nous l'avons déjà rappelé, la pratique établie est que le recenseur dresse la liste des logements dans le Registre des visites. Cette tâche est accomplie à l'occasion de la livraison des questionnaires. La combinaison de ces deux tâches contribue à une plus grande efficacité du fait que c'est l'occasion d'un contact personnel avec les recensés, que cela permet de déterminer la quantité de questionnaires à livrer dans les ménages nombreux et la langue des questionnaires, que le recenseur est en mesure de relever divers renseignements tels que le type de logement et, enfin, parce que cela favorise l'emploi d'une méthode d'échantillonnage particulièrement efficace. Cette méthode garantit l'actualité des listes puisque celles-ci sont établies très peu de temps avant le jour du recensement. On estime en outre qu'elle présente l'avantage d'accroître la motivation du recenseur puisque ce dernier est entièrement responsable du dénombrement du SD qui lui a été confié.

En dépit du succès de cette méthode dans le passé, il serait toutefois bon d'envisager la possibilité d'appliquer une nouvelle approche. En effet, nombre des avantages qu'offrait jusqu'à présent la livraison à domicile des questionnaires tendent à perdre de leur poids. En effet, le pourcentage de contacts établis à l'occasion de la livraison des questionnaires est de 50% et va en décroissant, 80% des ménages reçoivent actuellement le questionnaire bilingue et la taille des ménages diminue. Il devient de plus en plus difficile d'entrer dans certains types d'immeubles à appartements du fait que les occupants recherchent une sécurité accrue.

Il y a également d'importants changements au niveau du recrutement du personnel pour le recensement. Nous ne pouvons être certains que les recenseurs seront suffisamment motivés pour assurer une bonne couverture en 1991.

Le fait que le listage des logements ne soit pas automatisé signifie que les listes doivent être refaites entièrement à chaque recensement mais entraîne également certaines limites au niveau d'autres activités de recensement, qu'il s'agisse de la collecte ou du dépouillement des données, comme nous le verrons dans les paragraphes qui suivent.

L'autre approche possible, que nous exposerons en détail plus loin, consiste en un registre automatisé des adresses des ménages (RA). On y trouvera également une description d'un registre des adresses avec les applications possibles, un exposé des recherches accomplies dans ce domaine et une présentation des sources utilisables pour constituer un registre des adresses. Dans la conclusion du présent document, on trouvera une description du plan de recherche proposé pour le prochain recensement ainsi que divers points importants concernant l'établissement des registres des adresses.

Concept du registre des adresses et applications possibles

Le registre des adresses proposé ici consiste en un répertoire ordinolingue des adresses domiciliaires pour certaines régions ou pour l'ensemble du Canada. Chaque adresse doit correspondre à un logement, occupé ou non, selon la définition adoptée pour le recensement. Chaque adresse doit inclure les codes des éléments de la géographie du recensement à un niveau aussi détaillé que possible, incluant au moins le secteur de dénombrement. Chaque adresse peut également être assortie de certains renseignements complémentaires, comme le type de logement et, si possible, le numéro de téléphone. Cependant, le registre des adresses ne doit pas donner le nom des occupants du logement ni aucun autre renseignement personnel. Il doit être tenu à jour pendant les cinq années du cycle du recensement. Le fichier doit être modifié régulièrement et les ajouts, suppressions ou changements d'adresses doivent y être introduits aussitôt qu'ils sont connus.

Les principales utilisations d'un tel registre des adresses sont exposées ci-après:

Envoi des questionnaires par la poste

Dans la mesure où l'on peut utiliser un registre des adresses à jour, il est possible d'expédier les questionnaires par la poste au lieu de les livrer à domicile. Cette méthode est utilisée aux États-Unis pour une bonne partie de la population. Les résultats d'études de coûts comparatives effectuées dans le cadre de recherches antérieures font état d'économies de l'ordre de $40,000 à $652,000 dans le cas du recensement de 1981, et de $569,000 à $2.0 millions pour celui de 1986 (en dollars de 1976).

La majeure partie des économies réalisables vient de la suppression de la livraison des tâches du recenseur, y compris la suppression de la formation correspondante. Dans les secteurs de retour par la poste, du fait que le personnel n'aurait plus à s'occuper que du suivi, la taille du secteur attribué à chaque recenseur pourrait être plus importante, ce qui permettrait de réduire l'effectif sur le terrain. Cela pourrait également favoriser l'embauchage d'un personnel plus qualifié. En outre, un certain nombre de vérifications, telles que le contrôle postal et les vérifications sur le terrain des données géographiques effectués avant le début du recensement, pourraient être supprimées.

Amélioration de la couverture

Même si l'on n'opte pas pour le recensement postal, le registre des adresses pourrait être adopté en vue d'améliorer la couverture ou pour réduire les coûts du listage des logements, ou encore pour atteindre ces deux objectifs. On pourrait choisir d'utiliser le registre des adresses en remplacement du contrôle postal, devenu très onéreux ($1.4 million en 1981). Il s'agirait de remettre un Registre des visites de tout le secteur de dénombrement à chaque recenseur, qui n'aurait plus qu'à vérifier les adresses et à les mettre à jour. Cette façon de procéder permettrait de réduire les coûts et d'accroître la précision des données.

Délimitation des secteurs de tâche

Les secteurs de tâche attribués aux recenseurs (identiques ou équivalents aux SD existants) pourraient être délimités en fonction d'un nombre de ménages tirés des registres des adresses. Le fait de disposer de registres des adresses permettrait de réduire ou de supprimer le besoin de vérifications de la couverture géographique préalables au recensement et, probablement aussi, éviterait de fractionner certains SD au moment du dénombrement.

Emploi de nouvelles méthodes d'échantillonnage

Avec les registres des adresses, l'échantillon de questionnaires 2B pourrait être tiré à l'avance. Cela permettrait de mieux contrôler la sélection des unités de l'échantillon et, par conséquent, de réduire le biais consécutif à la livraison. Cela permettrait, en outre, l'emploi de méthodes plus complexes, notamment de varier le taux d'échantillonnage, de procéder à une stratification plus poussée ou de relier les échantillons. Une amélioration de la sélection de l'échantillon pourrait favoriser la réduction du taux d'échantillonnage et, par le fait même, entraîner une diminution du fardeau de réponse.

Traitement des logements collectifs

Si les logements collectifs sont indiqués dans les registres des adresses, des procédures spéciales peuvent être conçues à l'intention du personnel sur le terrain afin d'assurer que ces logements soient dénombrés adéquatement. En fait, il existe

déjà un registre des adresses pour les logements collectifs; il s'agit du fichier de listage des logements collectifs établi lors du dernier recensement. Un registre des adresses permettrait d'étendre la portée de ce fichier en facilitant la mise à jour durant la période intercensitaire.

Suivi téléphonique des cas de non-réponse totale

Pour le recensement de 1986, on prévoit que 80 à 90% des questionnaires seront retournés avant le début du suivi. Comme il y a 6 millions de ménages dans les secteurs de retour par la poste, cela veut dire qu'il faudra un suivi pour 600,000 à 1.2 million d'entre eux, soit l'équivalent de 10 à 20 enquêtes sur la population active.

Des économies considérables pourraient être réalisées si l'on procédait à un suivi téléphonique plutôt qu'à un suivi sur place dans le cas de ces ménages. En 1986, les recenseurs devront effectuer un suivi téléphonique auprès des ménages non répondants chaque fois que cela est possible. Toutefois, le recenseur ne peut connaître le numéro de téléphone que si le nom de la Personne 1 a été consigné au moment de la livraison, c'est-à-dire probablement dans 50% des cas.

Si le registre des adresses contient le numéro de téléphone correspondant à chaque adresse, le pourcentage précédent peut être grandement amélioré. Il serait alors possible d'établir un mini-répertoire téléphonique pour chaque SD à partir du registre des adresses. Un imprimé d'ordinateur répertoriant les logements par ordre d'adresse et donnant les numéros de téléphone permettrait au recenseur d'avoir le numéro de téléphone correspondant à tout logement qui a reçu un questionnaire mais ne l'a pas retourné.

Pour une telle application, il n'est pas essentiel que le registre des adresses soit complet ni absolument précis. Si une adresse ne figure pas dans le registre, le recenseur peut alors appliquer la procédure actuellement en vigueur.

Cette application doit être mise à l'essai en Ontario et au Québec à l'occasion du recensement de 1986, grâce à l'utilisation d'un fichier qui a déjà été acheté de la compagnie de téléphone. En intégrant le code postal à ce fichier, nous prévoyons restructurer ce dernier en fonction des SD de 1986 et fournir aux recenseurs les imprimés d'ordinateur dont nous venons de mentionner.

Centralisation du contrôle et du suivi téléphonique

Les arguments pour ou contre la centralisation du contrôle et du suivi téléphonique ayant été présentés dans les exposés précédents, nous ne reviendrons pas sur ce point. Néanmoins, si l'adoption d'une telle méthode devenait souhaitable, soulignons ici que le registre des adresses pourrait servir de base pour constituer un fichier destiné à contrôler le déroulement et les résultats d'une telle opération. Ce fichier de contrôle pourrait être utilisé:

a) pour enregistrer les questionnaires retournés par les recensés à leur arrivée;

b) pour consigner l'état d'avancement pour chaque questionnaire soumis au contrôle et au suivi;

c) pour établir les tâches des recenseurs chargés du suivi téléphonique et du suivi sur le terrain et contrôler l'avancement du travail, de façon automatique;

d) pour accomplir des tâches administratives telles que l'établissement des rapports du SIG, le calcul et la vérification de la rémunération;

e) comme source de données pour les analyses subséquentes se rapportant à cette opération.

En fait, le registre des adresses serait en quelque sorte une version automatisée du Registre des visites actuel.

Fichier de contrôle du traitement des données de recensement

Un fichier de contrôle basé sur le registre des adresses pourrait servir non seulement dans le cadre des activités de collecte des données mais également pour plusieurs étapes ultérieures du traitement. Ainsi, il pourrait être utilisé pour le contrôle du codage et pour la saisie des questionnaires. Il serait alors possible d'apparier directement aux questionnaires les données des Registres des visites (maintenant automatisés). Plusieurs étapes de l'opération actuelle de dépouillement au

bureau central pourraient être rationalisées par suite de l'automatisation des Registres des visites. Toute autre opération de recensement nécessitant l'accès aux Registres des visites serait également simplifiée.

À long terme, l'adoption d'un registre des adresses automatisé faciliterait l'intégration des opérations de collecte et de traitement des données. On pourrait prévoir un scénario selon lequel les questionnaires seraient envoyés par la poste puis saisis aussitôt qu'ils seraient revenus (après un prétraitement élémentaire, au besoin), vérifiés par ordinateur et mis en lots en vue du suivi. Étant donné que les questionnaires auraient été saisis, le suivi pourrait être fait par l'interview téléphonique assistée par ordinateur (ITAO). Une fois le suivi terminé, les questionnaires peuvent être soumis au contrôle et à l'imputation. Les chiffres provisoires de population peuvent être produits par la même occasion.

Application au fichier principal de la région (FPR)

Le registre des adresses peut être utilisé comme source de référence pour mettre à jour le fichier principal de la région. Il peut servir à étendre le niveau de codage du côté d'îlot au ménage, ce qui permettrait le géocodage direct des ménages pour n'importe quelle enquête. Cela faciliterait la mise en application de la politique relative à la nouvelle classification géographique type (CGT) et élargirait les possibilités de production de données régionales par Statistique Canada. Un niveau de codage plus détaillé du FPR serait également utile à bon nombre d'utilisateurs de ce fichier à l'extérieur du Bureau, notamment aux corps policiers municipaux.

Base de sondage pour les enquêtes-ménages

Outre les applications dans le cadre du recensement, un registre des adresses dûment à jour pourrait servir de base de sondage pour une grande variété d'enquêtes-ménages, comme l'enquête sur la population active, les enquêtes postcensitaires et intercensitaires en rapport avec le questionnaire de recensement ou les enquêtes-ménages indépendantes et ad hoc. Dans la mesure où il est possible de lier les enquêtes postcensitaires et intercensitaires à la base de données du recensement, il devient également possible de réduire le nombre de questions dans le questionnaire de recensement. Étant donné la demande accrue de données sur des sous-groupes de population particuliers, nous pourrions en venir à utiliser le recensement comme outil de sélection et à poser des questions plus détaillées dans une enquête de suivi. Le registre des adresses serait d'autant plus utile dans un tel contexte qu'il donnerait les numéros de téléphone.

Même si l'on ne se fonde pas sur les adresses pour tirer l'échantillon, le registre des adresses pourrait fournir des chiffres à jour sur les ménages en vue de l'établissement de plans d'échantillonnage d'enquêtes ad hoc ou de la mise à jour de plans d'enquêtes permanentes comme l'enquête sur la population active.

Résultats des recherches relatives au registre des adresses

Au milieu des années 1970, une série d'études de recherche a été effectuée concernant la faisabilité financière et opérationnelle de l'établissement et de la mise à jour d'un registre des adresses pour les régions urbaines du Canada. Un résumé de ces études est donné dans Booth (1976).

Le registre des adresses établi à cette occasion devait servir principalement dans le cadre d'un recensement de la population utilisant la méthode de retour par la poste. Les études ont donc porté plus particulièrement sur la couverture du registre des adresses. Trois mesures de la couverture ont été établies, sur la base de la couverture d'un registre des adresses idéal:

Intégralité %	= 100	–	nombre d'adresses valides ne figurant pas dans le RA	
			nombre d'adresses valides figurant dans un RA idéal	* 100
Précision %	= 100	–	nombre d'adresses non valides figurant dans le RA	
			nombre total d'adresses figurant dans le RA	* 100
Fiabilité %	= 100	–	(nombre d'adresses non valides figurant dans le RA + nombre d'adresses valides ne figurant pas dans le RA)	
			nombre d'adresses valides figurant dans un RA idéal	* 100

L'intégralité (%) est une évaluation du sous-dénombrement du RA, la précision (%) est une évaluation du surdénombrement du RA et la fiabilité (%) est une mesure combinée de la qualité.

Dans le cadre de la première étude, un registre des adresses a été établi pour cinq villes et tenu à jour pendant dix-huit mois à partir de données provenant de sources diverses, dont la Société canadienne d'hypothèques et de logement, les services octroyant les permis de construire et les bureaux de poste. Le tableau 1 donne les taux de cette étude relativement à l'intégralité (%).

Le principal problème, dans le cas du RA, concernait le repérage des sous-adresses dans le noyau des grandes régions urbaines. Les vérifications ont révélé que le sous-dénombrement était dû en grande partie au fait que de vieilles demeures avaient été converties en immeubles de deux ou trois appartements. Dans le cas des deux villes de moindre importance, toutefois, le RA a donné des résultats supérieurs aussi bien pour les maisons individuelles que pour les sous-adresses.

Le tableau 2 donne les taux de l'intégralité (%), de la précision (%) et de la fiabilité (%) pour les mêmes villes réparties en deux groupes.

TABLEAU 1. Intégralité % du registre des adresses (RA) par rapport au Registre des visites (RV) pour les cinq villes sélectionnées pour l'essai

	Type de logement					
Ville	Maison individuelle		Sous-adresse		Total	
	RA	RV	RA	RV	RA	RV
Montréal	99.48	99.19	98.08	98.42	99.19	99.03
Toronto	99.77	99.45	89.96	98.42	97.15	99.40
Vancouver	99.63	99.35	97.90	99.72	99.10	99.46
Sherbrooke	99.61	99.22	99.47	98.08	99.55	98.77
St. Catharines	99.65	99.62	98.50	97.57	99.34	99.06

TABLEAU 2. Intégralité, précision et fiabilité % du registre des adresses (RA) par rapport au Registre des visites (RV) pour les deux groupes des villes sélectionnées pour l'essai

	Montréal, Toronto et Vancouver		Sherbrooke et St. Catharines	
	RA	RV	RA	RV
Intégralité %	98.64	99.31	99.45	98.91
Précision %	98.78	96.78	98.17	98.72
Fiabilité %	97.43	95.97	97.59	97.62

D'après les résultats de la première étude, le RA semble offrir une qualité comparable à celle du listage effectué sur le terrain. La couverture du RA est moins bonne dans le cas des grandes régions urbaines alors qu'elle est meilleure pour celles de plus petite taille. Pour ce qui est de la précision et de la fiabilité, le RA est supérieur au RV dans le cas des grandes régions urbaines et le RV est légèrement supérieur pour celles de petite taille.

Les études subséquentes ont porté sur la possibilité d'établir et de mettre à jour un registre des adresses à partir d'une source unique, à savoir les feuilles relatives au lieu de livraison fournies par la Société canadienne des postes. Ces feuilles répertorient les adresses où le courrier est livré, pour tous les itinéraires de facteurs. Elles sont établies par le personnel du bureau de poste local et tenues à jour par le surveillant.

Une de ces études visait la ville de Trois-Rivières, qui faisait partie des villes sélectionnées pour l'essai de 1974. Le tableau ci-après donne les résultats obtenus à la fin d'une période de 11 mois pour un registre des adresses établi uniquement à partir des données fournies par la Société canadienne des postes.

Dans ce cas, les trois mesures pour le RA, c.-à-d. intégralité, précision et fiabilité sont supérieures à celles qui correspondent au RV. Le RV est supérieur dans un seul domaine, celui des logements collectifs. Des études semblables ont été effectuées pour quelques autres villes mais les résultats ne sont pas présentés ici. Les figures sont comparables.

En résumé, les études antérieures indiquent qu'il est possible d'établir et de tenir à jour un registre des adresses à partir d'une source de données unique (la Société canadienne des postes) et d'obtenir une qualité au moins équivalente à celle que donnerait un listage sur le terrain. Le seul point faible d'une telle méthode concerne l'identification des sous-adresses dans les grandes villes; il faudrait mettre au point des mesures permettant de corriger ce problème.

En dépit des possibilités qu'offre l'utilisation d'un registre des adresses, la recherche a été abandonnée parce que l'investissement que nécessite une telle méthode est apparu prohibitif.

Sources possibles pour un registre des adresses

Depuis que les études précitées ont été faites, les choses ont évolué de sorte qu'il pourrait devenir

TABLEAU 3. Intégralité, précision et fiabilité % du registre des adresses (RA) par rapport au Registre des visites (RV) pour Trois-Rivières

	Type de logement							
	Maison individuelle		Sous-adresse		Logement collectif		Total	
	RA	RV	RA	RV	RA	RV	RA	RV
Intégralité %	99.16	98.78	98.54	92.78	88.51	93.10	98.93	97.07
Précision %	99.10	99.42	96.88	90.06	100.00	100.00	98.47	96.79
Fiabilité %	98.26	98.21	95.36	82.54	88.51	93.10	97.40	93.79

possible, au cours des cinq prochaines années, d'utiliser l'ordinateur pour établir et tenir à jour un registre des adresses. Cette perspective sous-entend:

a) la disponibilité effective ou potentielle d'un nombre croissant de sources de données administratives à jour, telles que les registres des compagnies de téléphone ou des compagnies hydro-électriques, les fichiers des allocations familiales, de la sécurité de la vieillesse et de Revenu Canada;

b) l'adoption universelle du code postal pour ces fichiers qui, en combinaison avec le couplage du fichier principal des codes postaux et du fichier principal de la région, produit par la Division de la géographie, permet la restructuration automatisée de ces fichiers en fonction de la géographie du recensement;

c) l'amélioration des méthodes de couplage des enregistrements qui permet d'éliminer les doublets;

d) la réduction des coûts d'entreposage et de traitement résultant de l'utilisation d'importants fichiers de données.

Plus tôt dans l'année, un certain nombre de sources possibles d'adresses ont fait l'objet d'une recherche sommaire qui a abouti aux conclusions suivantes:

Feuilles relatives au lieu de livraison de la Société canadienne des postes

Contenu: Adresse, code postal, indicateur de local commercial ou résidentiel, nombre de boîtes aux lettres pour chaque adresse, autres indications de la charge de travail liée à la livraison du courrier.

Couverture: Par rapport aux résultats d'études menées dix ans plus tôt, cette méthode est équivalente au listage sur le terrain, sinon meilleure. Cette source ne donne que les adresses pour lesquelles le courrier est livré à domicile mais des renseignements complémentaires sont apparemment disponibles pour les autres adresses.

Disponibilité: Il existe un seul exemplaire de ces documents qui sont conservés dans les bureaux de poste. Il n'y a pas nécessairement de photocopieurs dans ces bureaux .

Fréquence: À déterminer.

Observations: La Société canadienne des postes est intéressée à entreposer ces listes sur support ordinolingue.

Les cartes illustrant les itinéraires des facteurs sont conservées dans les bureaux de poste.

Fichiers des compagnies de téléphone

Contenu: Nom, adresse et code postal de l'abonné, indicateur de local commercial ou résidentiel, numéro de téléphone.

Couverture: Les abonnés dont le numéro de téléphone n'est pas confidentiel.

Disponibilité: Bell Canada (pour la plus grande partie du Québec et de l'Ontario), sujet à négociation pour les autres régions.

Fréquence: Chaque trimestre.

Observations: Les renseignements relatifs à l'adresse sont fournis par le service de facturation et devraient donc être très précis.

On peut également avoir l'adresse où l'équipement est installé.

Listes des services d'évaluation municipaux

Contenu: Adresse, code postal, évaluation, catégorie (résidence régulière, chalet, ferme), type de logement.

Couverture: À déterminer (devrait être bonne en principe).

Disponibilité: Contre remboursement des coûts en Ontario, sujet à négociation pour les autres régions.

Fréquence: Chaque année.

Observations: L'évaluation établie par la municipalité peut être un sous-produit.

Compagnies hydro-électriques

Contenu: Adresse, code postal.

Couverture: Problèmes dus au fait que les immeubles à logements multiples ont généralement un compteur unique.

Disponibilité: Toutes les provinces, sauf l'Ontario, ont un service central de distribution de l'électricité; à déterminer.

Fréquence: Probablement chaque trimestre.

Observations: Les renseignements relatifs à l'adresse sont fournis par le service de facturation et devraient donc être très précis.

Allocations familiales

Contenu: Adresse, code postal.

Couverture: Les ménages avec enfants.

Disponibilité: Statistique Canada doit recevoir sous peu les fichiers de Santé et Bien-être social.

Fréquence: Probablement chaque mois.

Observations: Les renseignements relatifs à l'adresse sont ceux qui servent pour expédier les chèques, ils devraient donc être très précis.

Revenu Canada Impôt

Contenu: Adresse, code postal.

Couverture: Les contribuables.

Disponibilité: Actuellement à la disposition de Statistique Canada.

Fréquence: Chaque année.

Observations: Les données manquent d'actualité, il peut y avoir plusieurs contribuables pour la même adresse et certaines adresses peuvent correspondre à une exemption de taxe.

Fichier principal de la région

Le fichier principal de la région peut servir à établir un registre des adresses. Pour chaque côté d'îlot, le FPR donne le nom de la rue et la tranche d'adresses, par exemple, 2 à 54. On peut alors établir le registre des adresses correspondant en supposant qu'il y a un logement tous les deux numéros, par exemple, 2, 4, 6, ... 52, 54.

Contenu: Adresse, code postal, géographie du recensement.

Couverture: Identique à celle du FPR (c'est-à-dire la plupart des villes de 50,000 habitants et plus).

Disponibilité: Fichier de Statistique Canada.

Fréquence: Fichier permanent.

Observations: Le surdénombrement des adresses risque d'être un problème sérieux.

Base des appartements de l'enquête sur la population active

Contenu: Adresse de l'immeuble, code postal, géographie du recensement de 1981, nombre d'étages, nombre de logements.

Couverture: La plupart des RMR, tous les immeubles de 30 appartements et plus et de 5 étages et plus.

Disponibilité: Fichier de Statistique Canada.

Fréquence:	Fichier permanent.
Observations:	La mise à jour des données de la base peut remonter jusqu'à 12 mois.

Registres des visites établis lors du recensement de 1981

Contenu:	Adresse, indicateur d'exploitation agricole.
Couverture:	Les logements couverts par le recensement de 1981.
Disponibilité:	Sur papier seulement; devrait faire l'objet d'une saisie des données.
Fréquence:	Tous les cinq ans.
Observations:	Peut servir pour établir un registre des adresses; toutefois il faudrait bien entendu avoir recours à d'autres méthodes pour la mise à jour. Des études antérieures ont révélé que la saisie des données serait très onéreuse.

Outre les sources précitées, il existe plusieurs autres sources d'adresses qui mériteraient d'être étudiées, notamment les fichiers provinciaux des détenteurs de permis de conduire et de l'assurance-maladie, les sources commerciales comme celles qu'utilise le U.S. Bureau of the Census et les listes des électeurs établies par les directeurs fédéraux et provinciaux des élections.

Recherche à poursuivre et questions en suspens

Les étapes suivantes sont proposées en vue de poursuivre la recherche en la matière:

1. Il conviendrait d'entreprendre une étude plus détaillée des diverses sources et méthodes relativement à l'établissement et à la tenue d'un registre des adresses.

   Les diverses sources mentionnées auparavant devront être étudiées de façon plus approfondie de même que les méthodes de couplage des enregistrements permettant d'éviter les doublets. La recherche devra également englober l'étude d'un projet d'automatisation des listes des services postaux en coparticipation avec la Société canadienne des postes.

2. Il faudrait se tenir davantage au courant de l'expérience accumulée dans les autres pays qui se servent d'un registre des adresses et (ou) du recensement utilisant la méthode de retour par la poste.

3. Il importe de procéder à une actualisation des exercices de comparaison des coûts faits en 1976 de façon à avoir de meilleures estimations de la réduction possible des coûts grâce à l'utilisation combinée d'un registre des adresses et du recensement postal. Il conviendrait de procéder également à une comparaison des coûts pour l'enquête sur la population active.

4. Il faudrait établir un registre pilote en utilisant le couplage des enregistrements et des données administratives.

5. Il faudrait ensuite évaluer la qualité du registre pilote à l'occasion d'une enquête postcensitaire et de l'enquête sur la population active en cours. Cette évaluation devra porter à la fois sur la couverture du RA et sur la qualité des données qu'il contient. Une telle étude pourrait être intégrée aux autres essais sur le terrain relatifs au recensement de 1991.

6. Si le recensement postal s'avère faisable et laisse entrevoir des économies suffisantes, il faudra effectuer un essai sur le terrain afin d'évaluer la réaction des recensés et des interviewers, d'estimer les taux de réponse et de déceler les problèmes possibles.

7. Il conviendrait d'actualiser les comparaisons de coûts à partir des résultats de l'essai sur le terrain du recensement utilisant la méthode de retour par la poste.

8. Il faudrait procéder à un essai d'utilisation du RA comme base de sondage de l'enquête sur la population active.

Le tableau 4 présente un projet de calendrier des essais relatifs au prochain recensement.

Du fait qu'un registre des adresses pourrait être d'une grande utilité dans le cadre de nombreux programmes de Statistique Canada, il est proposé qu'un certain nombre de secteurs participent au travail. Les secteurs en question sont notamment le Programme de données régionales, la Division de l'exploitation des données administratives, la Division des méthodes de recensement et d'enquêtes-ménages (les deux sous-divisions), la Division des opérations des enquêtes, la Division de la géographie et la Division des services et du développement informatiques.

TABLEAU 4. Jalons de la recherche relative au registre des adresses	
Activité	**Date**
1. Étude de l'établissement et de la tenue du registre des adresses	Septembre 1986
2. Prise de connaissance de l'expérience accumulée dans les autres pays	Octobre 1986
3. Actualisation des comparaisons de coûts effectuées en 1976	Juin 1987
4. Établissement d'un registre pilote	Mars 1987
5. Essai postcensitaire portant sur la couverture du RA pilote	Juin 1987
6. Essai du recensement utilisant la méthode de retour par la poste	Juin 1988
7. Actualisation des comparaisons de coûts à partir des résultats de l'essai du recensement utilisant la méthode de retour par la poste	Mars 1989

Il faudrait en outre se pencher sur un certain nombre de points avant de mettre en oeuvre les applications précitées.

Tout d'abord, le recours au recensement utilisant la méthode de retour par la poste nous rend plus vulnérables aux erreurs commises par le personnel des postes de même qu'aux interruptions du service des postes et à l'augmentation du coût de ces services. Avec la méthode actuelle, par exemple, une grève des postes perturberait sérieusement la collecte des données mais ne l'empêcherait pas d'avoir lieu étant donné que les recenseurs pourraient passer reprendre les questionnaires si cela était nécessaire.

Deuxièmement, on ne sait pas comment le public réagirait à l'utilisation d'un registre des adresses. Il pourrait très bien assimiler le registre des adresses à un registre de la population et y voir une atteinte à la vie privée.

Troisièmement, un registre des adresses implique une normalisation géographique plus approfondie que celle qui existe actuellement. À l'heure actuelle, pour un même logement les données d'identification (district électoral, secteur de dénombrement et numéro de ménage) changent d'un recensement à un autre. Il conviendrait d'adopter une forme de programme d'îlots en vue de préserver la stabilité des données d'identification géographique jusqu'au niveau du côté d'îlot.

Quatrièmement, il faut déterminer la portée du registre des adresses. Pour le moment, un tel registre semble plus facilement applicable aux régions urbaines qu'aux régions rurales. Or c'est dans ces dernières qu'il y aurait le plus d'économies à réaliser en éliminant les frais de déplacement liés à la livraison des questionnaires.

Cinquièmement, il faudrait savoir dans quelle mesure les diverses sources de renseignements dont on a besoin pour tenir un registre des adresses sont fiables. Ainsi, il y a peu de temps, la Société canadienne des postes a entrepris de réduire les services de livraison du courrier à domicile afin de réduire ses coûts. Cela veut dire que les adresses en question pourraient ne plus être disponibles ou être difficiles à obtenir.

Enfin, il faut se demander s'il est possible de réduire le risque inhérent au remplacement systématique de l'approche traditionnelle par une nouvelle approche basée sur l'utilisation du registre des adresses. Cela revient, en fait, à passer de l'approche fondée sur l'échantillon aréolaire à une approche fondée sur une liste. La recherche effectuée, pour encourageante qu'elle soit, a porté sur un nombre relativement restreint de villes.

Une solution qui pourrait être envisagée pour 1991 consisterait à établir et à tenir à jour un registre des adresses à l'échelle nationale sans toutefois l'utiliser dans le cadre de la collecte des données. La qualité (intégralité, précision et fiabilité) d'un tel registre serait évaluée aussitôt le recensement de 1991 terminé, au moyen d'une comparaison intégrale des données et d'une vérification sur le terrain. Comme la comparaison porterait sur deux listes établies indépendamment l'une de l'autre, l'évaluation de la qualité du RA serait vraisemblablement fiable. Toutefois, nous n'aurions aucune idée des problèmes que pourrait poser l'utilisation d'un RA sur le terrain. Une comparaison portant sur cet aspect particulier serait difficile et onéreuse à effectuer.

Une seconde approche pourrait être envisagée qui consisterait à conserver la méthode de collecte

actuelle (livraison des questionnaires à domicile) mais à remettre aux recenseurs une copie du registre des adresses correspondant à leur SD que ceux-ci utiliseraient comme point de départ pour faire le listage des logements. Les mises à jour effectuées par les recenseurs seraient ensuite saisies et la qualité du RA serait alors évaluée en comparant le RA original à la version révisée. Cela reviendrait à intégrer la vérification sur le terrain du RA au recensement de 1991. Cependant, étant donné que la vérification du RA serait une vérification dépendante, il y aurait un risque que la qualité du RA soit surestimée. Par contre, nous aurions une bonne idée de ce que donne l'utilisation du RA sur le terrain et nous pourrions peut-être réaliser des économies.

Quelle que soit la méthode retenue, il est clair qu'une exploitation optimale d'un RA n'est possible que si l'on opte délibérément pour le recensement utilisant la méthode de retour par la poste. Un tel changement doit avoir été soigneusement essayé et soupesé, mais le fait que cette méthode soit utilisée avec succès aux États-Unis est une constatation encourageante. C'est à Statistique Canada de décider s'il convient de faire le pas à l'occasion du recensement de 1991.

BIBLIOGRAPHIE

Booth, J.K., 1976. "A Summary Report of All Address Register Studies Completed to Date", Statistique Canada, rapport E-414E.

Booth, J.K., 1976. "NAR cost comparisons - Present Worth Approach", Statistique Canada, note de service interne.

Résumé de la période de questions

À l'occasion de la période de questions et de discussion divers points et questions ont été soulevés.

Couverture

La première préoccupation a été de déterminer si une comparaison internationale des résultats des évaluations de la couverture permet de mettre en évidence les raisons du sous-dénombrement, par exemple, de préciser les types de régions ou de population pour lesquels les taux de sous-dénombrement sont particulièrement élevés. Des tentatives de comparaison ont été faites mais elles n'ont abouti à rien parce que les techniques utilisées dans différents pays pour mesurer l'erreur de couverture étaient trop différentes. Ainsi, aux États-Unis les estimations obtenues à partir de l'analyse démographique et des enquêtes postcensitaires diffèrent énormément. La seule façon d'éviter ces différences dues aux méthodes serait de procéder à une expérience contrôlée.

Il a également été souligné qu'il peut être plus approprié d'utiliser des techniques différentes pour certaines populations ou pour certains niveaux d'agrégation.

La troisième constatation faite est que les États-Unis et le Canada apportent un soin particulier à la mesure de l'erreur de couverture alors que le Royaume-Uni s'applique à mesurer l'erreur de réponse, encore que ces trois pays évaluent les deux types d'erreur.

Enfin la dernière question soulevée, de nature plus générale, consistait à déterminer si les compromis entre la qualité et l'actualité des données étaient différents selon qu'il s'agissait de recensements, dont les données sont utilisées pendant cinq ou dix ans, ou d'enquêtes mensuelles, trimestrielles et annuelles. On a fait valoir qu'il est impossible, en pratique, de prendre davantage de temps au cours du processus de recensement pour améliorer partiellement la qualité des données et que l'actualité des résultats du recensement est un facteur important car les caractéristiques d'une population peuvent changer rapidement.

Contrôle et imputation

En ce qui concerne le contrôle et l'imputation, la discussion a porté principalement sur le fait que les organismes statistiques peuvent avoir à rendre compte des méthodes qu'ils utilisent pour contrôler ou imputer les données. En outre, on note une préoccupation grandissante chez les utilisateurs vis-à-vis de ces méthodes. Les exemples fournis à ce titre englobaient les données sur la langue dans le cas du recensement du Canada et celles sur la race et l'origine hispanique dans le cas du recensement des États-Unis. Un certain nombre de possibilités ont été envisagées pour l'avenir. Les organismes statistiques pourraient avoir à négocier les modalités du contrôle et de l'imputation des données tout comme ils négocient déjà le libellé des questions. Nous pourrions décider de ne plus imputer les données avec toutes les implications qui en résulteraient pour les diverses séries de données tirées de mêmes questions. Il faut également tenir compte de l'importance pour bon nombre d'utilisateurs de pouvoir comparer les données chronologiques.

La discussion a également porté sur la possibilité de publier des données non vérifiées dans un premier temps et des données dûment vérifiées ultérieurement. Certains ont fait valoir que cette solution risquait de retarder la parution des données finales. La question de savoir s'il vaut mieux avoir plus rapidement des données non vérifiées au prix d'un retard de la parution des données finales reste en suspens.

Registres des adresses

Plusieurs questions ont été posées relativement à la portée du type de registre des adresses envisagé. On a fait allusion à la possibilité d'y inclure les exploitations agricoles en utilisant le registre des fermes comme source de référence. Il est apparu que, pour 1991, l'utilisation des registres des adresses devrait plutôt être restreinte aux régions urbaines.

Les représentants des États-Unis ont évoqué un certain nombre de raisons sur lesquelles était fondée l'adoption du registre des adresses et de la méthode d'envoi et de retour des questionnaires par la poste. Ils ont fait valoir que bien que les registres des adresses soient onéreux, les États-Unis avaient opté pour cette solution parce qu'elle permet une amélioration de la couverture et du contrôle. En outre, l'importance de la population des États-Unis impose le recours au recensement utilisant la méthode de retour par la poste du fait qu'il est impossible que tous les logements soient dénombrés par des recenseurs. Troisièmement, les États-Unis jouissent de l'avantage que, pour de nombreuses régions, des tests, disponibles à bon prix et de bonne qualité, ont été effectués par les fournisseurs des registres; il n'en va pas nécessairement de même au Canada. Quatrièmement, si le coût des registres des adresses est élevé, il reste qu'il peut être amorti sur plusieurs recensements. En dépit de ce qui précède, il a été jugé intéressant pour Statistique Canada de poursuivre ses recherches relativement à l'utilisation de registres des adresses.

À la question portant sur l'intégration éventuelle des registres des adresses au système TIGER pour 1990, les représentants des États-Unis ont répondu par la négative. La raison qui a été donnée est que le système TIGER est un projet colossal en lui-même et que le fait d'y ajouter une étape supplémentaire serait risqué à ce stade-ci.

La question du respect de la vie privée a également été abordée. Pour être utile un registre des adresses doit donner les numéros de téléphone. Or, il n'en faudrait pas moins pour que les registres des adresses deviennent des registres de la population.

Des recherches doivent être entreprises pour établir dans quelle mesure le public est prêt à accepter les registres des adresses. On a souligné qu'aux États-Unis les registres des adresses sont traités comme des documents confidentiels.

Enfin, il a été suggéré que la mise à l'essai des registres des adresses en 1991 soit soumise à une troisième approbation. Il s'agirait de choisir une région cible où une méthode combinant l'établissement des registres des adresses et l'envoi et le retour des questionnaires par la poste serait mise à l'essai; il y aurait un plan d'action prévisionnel en cas d'échec de l'essai.

SÉANCE: RÔLE DE LA RECHERCHE ET DES ESSAIS

Président: E. T. Pryor
Directeur général
Statistique démographique et recensement

Vendredi, le 11 octobre 1985

PLANIFICATION DU RECENSEMENT DE 1991 AU ROYAUME-UNI

DAVID PEARCE

OFFICE OF POPULATION CENSUSES AND SURVEYS ROYAUME-UNI

Introduction

Avant de commencer, je tiens à vous dire à quel point les trois dernières journées m'ont été agréables. La rencontre de spécialistes d'autres pays dans le domaine du recensement a été pour moi une expérience très stimulante. Brièvement, je voudrais vous faire part de quelques-uns des aspects de la conférence dont j'ai tiré profit.

Premièrement, il y a la question de l'échéancier prévu pour la planification du recensement de 1991 au Royaume-Uni. Deuxièmement, la conférence m'a rassuré sur certains points en ce qu'elle m'a permis de constater que d'autres pays éprouvaient les mêmes difficultés que nous. Troisièmement, la conférence a confirmé le fait que bien que la nouvelle technologie offre des possibilités très intéressantes que nous sommes impatients d'exploiter, nous ne pouvons y recourir sans tenir compte de facteurs tels que les coûts, l'actualité et la précision. Enfin, il importe non seulement de veiller à ce que les questions soient bien conçues d'un point de vue technique mais aussi de nous assurer qu'elles soient acceptées du public.

Pour vous donner un aperçu de la façon dont les recensements sont effectués au Royaume-Uni, je voudrais vous parler de quelques-unes des leçons que nous avons tirées de la planification des recensements, depuis celui de 1971 jusqu'à celui de 1991. Ensuite, je vous expliquerai brièvement où nous en sommes en ce moment et la voie dans laquelle nous nous sommes engagés. Je n'aborderai pas ici de questions d'intérêt général, par exemple, la nécessité de continuer à nous assurer la collaboration du public et à mériter sa confiance en ce qui concerne la confidentialité des données et la nécessité de trouver un juste compromis entre les besoins des utilisateurs et l'exploitation des moyens technologiques qui s'offrent à nous. J'écarte donc ces questions pour m'attacher à des questions plus précises.

Expériences des recensements de 1971 à 1991

En 1971, nous avons eu des difficultés parce que le programme utilisé pour faire les totalisations manquait de souplesse, était inefficace et était dispendieux en temps d'ordinateur. C'était un très gros programme, peu coordonné et qui est devenu presque impossible à contrôler.

En 1981, nous avons essayé d'améliorer le programme de totalisations en utilisant des flux de données ce qui a permis de créer des séries de données. Ensuite, nous avons établi des dates cibles pour la production des données requises aux fins de l'application des programmes de subventions administrés par les autorités locales. Enfin, nous avons établi un échéancier et défini nos priorités en ce qui concerne la diffusion des résultats.

Diffusion des résultats du recensement de 1981

Dans quelle mesure avons-nous atteint nos objectifs en 1981? Premièrement, les données nécessaires à l'application des programmes de subventions ont été produites à temps. Il s'agit là d'une réalisation digne de mention. Également, entre octobre 1981 et juin 1982, un bulletin contenant les résultats provisoires du recensement a été distribué dans chaque comté. Nous avons aussi amélioré le programme de publications par rapport à 1971: l'ensemble des produits ont paru entre octobre 1981 et juin 1982 et la publication sur le pays de naissance a paru en février 1983. Pour le recensement de 1971, les publications avaient été diffusées de mai 1972 à novembre 1973 et la publication sur le pays de naissance avait paru en décembre 1974. Les délais de diffusion ont donc été beaucoup plus courts pour le recensement de 1981.

Toutefois, bien que le programme de diffusion des résultats du recensement de 1981 ait été une réussite, il reste encore des points à améliorer pour le prochain recensement.

Méthodes de collecte

Une des questions qu'il faut examiner de plus près est celle du sous-dénombrement. Bien que le taux moyen de sous-dénombrement ait été de 0.5% seulement, il a beaucoup varié d'un endroit à l'autre. Par exemple, dans les régions de l'intérieur, il était de 2.5%. Les principales causes du sous-dénombrement étaient les suivantes: premièrement, les personnes oubliées dans les ménages et, deuxièmement, les ménages codés "absents" le soir du recensement quand, en réalité, au moins une personne occupait le logement. Les résultats d'une enquête effectuée après le recensement ont révélé que le pourcentage des

ménages absents a été beaucoup moins grand que le pourcentage établi à partir des données fournies par les recenseurs. Il faudrait donc revoir la méthode de collecte afin de déterminer s'il y a lieu de modifier les critères sur lesquels les recenseurs doivent se fonder et les mesures de vérification appliquées au travail de ces derniers.

Qualité des réponses

Au Royaume-Uni, nous faisons appel à des procédures complexes pour vérifier la qualité des réponses. Le taux d'erreur de réponse pour certaines questions indique que ces dernières ont été mal comprises. Par exemple, le taux d'erreur pour la question sur le nombre de pièces dans le logement était de presque 30% et, pour la question sur la situation du recensé vis-à-vis de l'emploi, d'environ 10%. Dans l'ensemble, le public et les utilisateurs semblent prêts à accepter que les données ne soient pas d'une exactitude absolue. Toutefois, il faut quand même établir le taux d'erreur de réponse pour chaque question afin d'évaluer la pertinence des données.

Bases de données sur la population

En 1981, les chiffres de population provisoires ont été établis à partir du registre du recenseur qui représente un compte manuel de la population (le compte provisoire était de 49 millions d'habitants tandis que le compte final était de 49.15 millions).

Ensuite, on a créé deux bases de données sur les résidents habituels. La première contenait le compte des résidents habituels "présents ou absents" mais non les ménages dont tous les membres étaient absents le soir du recensement. Le compte y était donc moins élevé (48.52 millions) que dans la base de résidents transférés qui comprenait les ménages absents (49.15 millions).

État d'avancement des travaux de planification du recensement de 1991

Nous avons défini un certain nombre de lignes directrices pour la planification du recensement de 1991. Dans un premier temps, il s'agissait de créer un comité composé des directeurs adjoints et de représentants des divisions intéressées. Ce comité a déjà défini les principales modalités du recensement de 1991. Il a décidé, premièrement, que le recensement de 1991 sera un recensement simple et complet. Deuxièmement, le processus de consultation doit commencer dès maintenant. Troisièmement, il faut réduire les délais de diffusion des résultats. Quatrièmement, les mesures de sécurité prises en 1981 pour assurer la confidentialité des données doivent être maintenues.

De plus, nous nous sommes entendus sur d'autres aspects tels que le calendrier de consultation. Également nous avons décidé que l'on devait mettre au point un programme d'essais qui, en 1987, comportera un test pour l'ensemble du questionnaire et un test distinct pour la question sur l'origine ethnique. En 1987 encore, nous prévoyons produire un livre blanc sur la planification du recensement de 1991, qui sera présenté au Parlement. Ce livre devra aussi être disponible au public.

Création d'équipes de projet

La création d'équipes de projet permet l'identification de questions spécifiques qui devront éventuellement faire l'objet de recherches et être incluses dans le programme d'essais. Jusqu'à maintenant six équipes ont été constituées. La première équipe est chargée d'étudier la question de la géographie en ce qui a trait aux procédures de dénombrement. En Angleterre, à la fin des années 1970, les unités postales cartographiées représentaient approximativement 1.6 million de codes postaux. Chaque unité postale est composée de 15 à 17 adresses. La numérotisation des unités postales bénéficierait non seulement à la planification des districts de dénombrement mais aussi au programme de diffusion. Qui plus est, un nombre important de données statistiques est relié aux unités de codes postaux. La possibilité de dériver des données statistiques de différentes sources compatibles géographiquement faciliterait certaines implications. Enfin, il ne fait aucun doute que de nombreux utilisateurs seraient intéressés à obtenir des statistiques produites en fonction du code postal. Le problème toutefois est de trouver les fonds nécessaires pour effectuer des recherches à ce sujet. En effet, il est difficile pour les planificateurs de recensement d'obtenir des fonds pour un programme de recherche et d'essais, particulièrement lorsque les plans pour ce recensement n'en sont qu'à leur stade initial de développement.

Le deuxième groupe étudie les systèmes de traitement qui devraient être utilisés. Un des principaux problèmes tient à l'étendue du district. Le choix du matériel informatique soulève également des difficultés. Je pense notamment au problème posé par l'acquisition du matériel et par la durée des essais. De plus, avant d'arrêter un choix, il faudra trouver des réponses à un certain nombre de questions: de quel genre de système avons-nous besoin? Faut-il choisir un système double? Faut-il choisir un système de contrôle centralisé ou décentralisé? La saisie des chiffres de base devrait-elle être effectuée au moyen d'un lecteur optique de caractères ou plutôt d'un lecteur

optique de marques, et la saisie des autres chiffres, par saisie au clavier avec programme de contrôle? Quelles opérations manuelles faut-il informatiser? Voilà les principales questions auxquelles il faudra trouver une réponse dans un proche avenir.

Le troisième groupe a pour tâche de déterminer s'il est possible de faire l'essai de questions relativement à la variable origine ethnique/langue maternelle. En 1981, nous n'avons pas inclus cette variable dans le questionnaire mais nous nous sommes engagés à en faire l'essai pour le recensement de 1991. Afin d'inclure ces questions dans le prochain recensement, nous devons d'abord démontrer qu'elles permettront d'obtenir les résultats désirés, les formuler d'une manière qui soit acceptable et montrer qu'elles ont une raison d'être dans un recensement.

Le quatrième groupe est chargé d'établir des bases de données sur la population. La principale question, en ce qui concerne ces bases, concerne les résidents habituels. Deux possibilités sont à envisager: compter les visiteurs dans leur région de résidence habituelle ou obtenir des données pour les ménages dont tous les membres étaient absents, ce qui serait probablement plus facile. Quoi qu'il en soit, le fait de remplir deux formules pose des problèmes juridiques dont nous devons tenir compte.

Le cinquième groupe est chargé d'examiner les procédures de dénombrement mais je n'examinerai pas son mandat.

Enfin, le sixième groupe étudiera les activités économiques. Premièrement, nous devons mettre à jour la classification des professions. Deuxièmement, nous devons nous assurer que la mise à jour rend compte des changements survenus dans le domaine de la profession depuis dix ans. Troisièmement, la représentation des femmes dans le domaine de la profession a souvent été critiquée et mériterait qu'on l'examine de plus près. Quatrièmement, nous devons faire la mise à jour en tenant compte des classifications internationales des professions et des autres classifications utilisées au pays. Enfin, nous devons réévaluer les critères utilisés pour regrouper les classifications.

Je vous remercie de votre attention.

QUESTIONS IMPORTANTES DU RECENSEMENT DE LA POPULATION ET DU LOGEMENT DE 1990 AUX ÉTATS-UNIS

PETER A. BOUNPANE

DEMOGRAPHIC CENSUSES
U.S. BUREAU OF THE CENSUS

Introduction

Aux États-Unis, le prochain recensement de la population et du logement aura lieu le 1er avril 1990 et marquera le 200e anniversaire de la tenue de recensements au pays. Bien qu'il reste encore quatre ans et demi avant ce recensement, les travaux de planification ont commencé il y a déjà un certain temps. Compte tenu du nombre de décisions qui doivent être prises et de la longueur des délais d'exécution, il est nécessaire d'entreprendre les préparatifs bien à l'avance.

Comme le prochain recensement marquera le bicentenaire de la tenue de recensements aux États-Unis, il a une importance particulière. Nous voulons que ce recensement soit digne de notre longue tradition. Le recensement de 1990 est le 21e d'une chaîne qui, depuis 1790, n'a jamais été rompue. Les données qui y seront recueillies nous mèneront jusqu'au début du XXIe siècle.

Progresser, c'est bâtir sur les expériences du passé, positives et négatives. Dans un premier temps, nous avons donc procédé à une évaluation approfondie du recensement de 1980. Dans l'ensemble, le recensement de 1980 a été une réussite. Plusieurs réalisations méritent d'être signalées:

- D'après les estimations, nous avons obtenu une meilleure couverture en 1980 qu'en 1970.
- Nous avons réussi à produire les chiffres de population requis pour procéder au redécoupage des cartes électorales dans les délais fixés par la loi.
- La campagne d'information et de promotion a très bien fonctionné; nous avons obtenu un taux de réponses postales de 83% et une bonne couverture.
- Nous avons produit davantage de données, en particulier sur les petites régions.

Cela ne veut cependant pas dire que nous n'avons pas eu de problèmes au recensement de 1980. Voici les principales difficultés auxquelles nous avons dû faire face:

- La diffusion de certains produits a été retardée, particulièrement en ce qui concerne les produits contenant les données recueillies avec le questionnaire-échantillon ou questionnaire complet.
- Les cartes et d'autres documents relatifs au découpage du territoire ont été produits séparément, en différentes opérations manuelles, ce qui a entraîné des retards et des inexactitudes que nous avons dû corriger avant de pouvoir diffuser les produits.
- Bon nombre des bureaux temporaires ont eu de la difficulté à recruter du personnel et à le garder jusqu'à la fin des opérations.
- Certaines questions ont été soulevées au sujet de la fidélité des données du recensement de 1980. Ces questions ont quelquefois conduit à des litiges légaux. C'est pour cette raison que certains bureaux temporaires ont dû rester ouverts plus longtemps que prévu.

Au cours des deux prochaines années, nous devrons faire des choix très importants. Dans la plupart des cas, il n'y aura pas de solution unique et idéale; nous devrons donc nous efforcer de définir le meilleur compromis possible. Pour arrêter ces choix, nous avons établi les critères que voici:

- Premièrement, en vertu de la Constitution et de la loi, nous devons remettre au président les chiffres de population requis pour le redécoupage de la carte électorale fédérale au plus tard le 31 décembre 1990. Les chiffres requis pour la redistribution des districts par les États doivent être déposés au plus tard le 1er avril 1991. Tout changement apporté au recensement de 1990 devra aider le Census Bureau à respecter ces délais.
- Deuxièmement, nous voulons diffuser les données du recensement de 1990 plus rapidement que jamais.
- Troisièmement, nous voulons que le coût global du recensement soit raisonnable. Notre objectif pour 1990 est de maintenir le coût à

l'unité (rajusté pour tenir compte de l'inflation) au niveau de 1980.

- Quatrièmement, nous visons la plus grande exactitude possible des données. Il s'agit là d'un objectif de taille car nous voulons maintenir les coûts à un niveau raisonnable et accélérer les opérations sans toutefois compromettre la qualité des résultats.

- Cinquièmement, en ce qui concerne le choix des questions, nous devons trouver le plus juste compromis possible entre les besoins en données et le fardeau de réponse. Les demandes légitimes de données seront plus nombreuses mais nous devrons veiller à ce que le questionnaire ne soit pas trop long tout en s'assurant que les données de base seront recueillies.

- Sixièmement, nous devons assurer la plus stricte confidentialité des réponses de chaque recensé. La réussite du recensement dépend directement de la participation du public, et celui-ci ne collaborera que s'il a confiance en nous.

Nous allons maintenant parler des principaux points à considérer dans la planification du recensement de 1990: procédures de base, automatisation, personnel, campagne d'information et de promotion ainsi que le contenu du questionnaire.

Procédures de base

Examinons d'abord les procédures de base que nous utiliserons au recensement de 1990. Comme en 1980, nous prévoyons utiliser la méthode de l'envoi et de retour par la poste pour presque toutes les régions du pays. Nous enverrons un questionnaire à tous les ménages inscrits sur notre liste d'adresses (nous ferons tous les efforts voulus pour nous assurer que la liste est aussi complète que possible) et nous leur demanderons de retourner par la poste le questionnaire rempli. Nous ne communiquerons qu'avec les ménages qui n'auront pas retourné le questionnaire ou pour lesquels nous aurons besoin de précisions. Dans les régions peu peuplées, où la méthode de retour par la poste est difficilement utilisable, des recenseurs rendront visite à chaque ménage. Nous devons pour toutes les régions établir des procédures pour effectuer le dénombrement des logements collectifs (par exemple, les résidences universitaires, les camps militaires et les prisons), nous devons aussi prévoir des mesures de contrôle de la qualité et des procédures d'amélioration de la couverture afin d'assurer la plus grande exactitude possible du recensement.

Pour le recensement de 1990, nous étudions la possibilité de modifier ou d'améliorer certaines des procédures utilisées en 1980. Par exemple, dans le recensement d'essai que nous avons fait en 1985 à Jersey City (New Jersey), nous avons procédé en deux temps: nous avons d'abord recueilli les données de base auprès de l'ensemble de la population et, deux mois plus tard, les données-échantillon auprès d'un échantillon de répondants. En 1980, toutes les données ont été recueillies en une seule étape. Certaines personnes croyaient qu'en procédant en deux étapes on obtiendrait peut-être davantage de données dans les régions urbaines où le dénombrement est difficile à effectuer. Toutefois, les avantages attendus de cette méthode ne se sont jamais concrétisés. En fait, le taux de retour par la poste pour la seconde étape a été si bas (15%) que nous avons renoncé à effectuer des suivis. Nous n'avons pas l'intention de reprendre cet essai d'ici 1990.

Dans le recensement d'essai effectué à Tampa (Floride) en 1985, nous avons fait l'essai de fiches de rappel. Nous voulions déterminer si nous pouvions améliorer le taux de retour par la poste et réduire les coûts liés aux visites de rappel. Quelques jours après l'envoi des questionnaires, nous avons fait parvenir une fiche de rappel aux recensés qui n'avaient pas retourné leur questionnaire. L'essai a révélé que ces fiches peuvent réduire les coûts et nous entendons nous en servir de nouveau dans nos essais de 1986.

Nous continuerons à perfectionner nos procédures dans les essais que nous effectuerons en 1986. Dans le recensement d'essai qui aura lieu dans une partie du comté de Los Angeles en Californie, nous essaierons de trouver des moyens de corriger les problèmes que pose la livraison des questionnaires dans certains immeubles à logements multiples. Dans un autre recensement d'essai que nous effectuerons dans plusieurs comtés ruraux du East Central Mississippi, nous testerons également diverses méthodes précises de livraison des questionnaires aux ménages situés dans des zones de livraison des routes rurales. L'une d'elles consistera à confier la livraison à des recenseurs plutôt qu'au service postal. Les recenseurs profiteront de l'occasion pour mettre à jour, au besoin, notre liste d'adresses. Dans cet essai, nous réévaluerons également certaines de nos procédures de dénombrement dans les réserves indiennes du Mississippi.

Dans tout recensement, il importe de prévoir des procédures pour améliorer la couverture. En 1980, par exemple, nous avons pris plusieurs mesures pour accroître l'exactitude de notre liste

d'adresses, nous avons fait une vérification des logements "vacants" dans le but de vérifier s'ils étaient occupés, nous avons contrôlé nos listes d'adresses en les comparant à d'autres, nous avons fait une campagne conçue pour sensibiliser les gens à la nécessité de se faire recenser, etc. Pour le recensement de 1990, nous allons évaluer les procédures d'amélioration de la couverture utilisées en 1980 pour déterminer celles que nous devons conserver et pour définir les perfectionnements possibles.

Un des programmes d'amélioration de la couverture auquel nous accordons une attention particulière est le Local Review Program. En 1980, nous avons pour la première fois demandé aux autorités locales de 39,000 entités légales de vérifier nos chiffres de population avant la fermeture de nos bureaux temporaires. Les autorités locales ont noté tout écart entre nos chiffres et les leurs et nous avons pu effectuer les corrections qui s'imposaient. Nous voulons améliorer ce programme pour le recensement de 1990. Nous sommes en train de mettre au point un programme conçu pour que les autorités locales puissent vérifier le nombre d'adresses avant le jour du recensement et les chiffres actuels de population, avant la fermeture des bureaux temporaires. Les autorités locales auront plus de temps pour préparer leurs données et pour revoir nos chiffres. Nous entendons entrer en communication avec les autorités locales plus rapidement que nous ne l'avons fait en 1980 et nous envisagerons la possibilité d'organiser des séances de formation, en collaboration avec des organismes des États, pour les aider à se préparer à appliquer le programme.

En plus de chercher à améliorer nos procédures pour accroître l'exactitude du recensement, nous continuerons d'examiner diverses techniques d'évaluation du sous-dénombrement et de correction des chiffres de population. Les méthodes d'ajustement des chiffres de population doivent être non seulement valables sur le plan statistique mais aussi acceptables sur le plan politique et juridique. Les méthodes choisies doivent également être faciles à appliquer et permettre de respecter les délais imposés par la loi. Le recensement d'essai de 1986 dans le comté de Los Angeles nous permettra de déterminer s'il est possible d'évaluer rapidement la couverture de manière à pouvoir, au besoin, corriger les chiffres dans les délais imposés.

Automatisation

Un autre choix que nous devons faire, concerne l'automatisation. Je passerai très rapidement sur cette question puisqu'elle a été examinée de manière plus approfondie dans d'autres documents pour cette conférence.

En gros, notre but est d'automatiser davantage le recensement de 1990 afin d'accélérer le travail, de réduire au maximum les dépenses et d'accroître la qualité des résultats. Dans le passé, la collecte des données et une bonne partie du dépouillement se faisaient à la main. L'automatisation peut nous aider à supprimer une montagne de papiers et des milliers d'opérations manuelles et à effectuer l'ensemble des opérations du recensement avec beaucoup plus d'efficacité et de contrôle.

Il y a deux changements à apporter pour automatiser davantage le recensement de 1990. Le premier est l'automatisation de bon nombre de tâches qui, dans le recensement de 1980 et dans les recensements précédents, étaient accomplies manuellement. Soulignons notamment la production de cartes géographiques, la mise à jour des listes d'adresses, la vérification et le contrôle des questionnaires au moment de leur réception, le codage des réponses écrites sur les questionnaires et l'établissement des rapports sur les coûts et l'état d'avancement des travaux.

Le deuxième changement consiste à commencer l'étape du dépouillement automatisé des données plus tôt qu'en 1980. Lors du recensement de 1981, la collecte et le traitement étaient deux étapes bien distinctes. Nous ne commencions le traitement qu'une fois terminé tout le travail dans un bureau de district (habituellement 5 à 7 mois après le jour du recensement). En 1990, nous commencerons le traitement simultanément avec la collecte.

Une fois prise la décision de commencer le traitement des données plus tôt, il reste à déterminer où (c'est-à-dire centralisé dans quelques endroits ou décentralisé dans plusieurs endroits) et comment il sera effectué (c'est-à-dire en utilisant la conversion microfilm/bande magnétique, l'introduction par clavier ou le lecteur optique de marques). Nous devons prendre des décisions au plus tard en septembre 1986 en prévision du long processus d'acquisition du matériel. Mais nous essaierons de régler le plus de problèmes possible avant cette échéance, à l'occasion d'une réunion qui doit avoir lieu à la mi-octobre 1985.

Recrutement du personnel

Bien que la conception d'un système efficace de recensement soit importante, encore faut-il avoir un personnel compétent capable de mener la tâche à bonne fin. En 1980, dans certaines régions, nous avons eu de la difficulté à embaucher du personnel compétent et à le garder. Cet état de choses tenait

en partie au fait que dans certaines régions le taux de rémunération que nous offrions aux employés n'était pas concurrentiel, à la nature temporaire des emplois et au fait que certaines tâches, notamment les visites chez les non-répondants, peuvent être très difficiles. Nous cherchons actuellement de nouveaux moyens de recruter, d'embaucher et de garder des employés temporaires pour le recensement de 1990.

Dans le recensement d'essai que nous avons effectué à Jersey City en 1985, nous avons engagé un certain nombre d'employés à temps partiel. En 1980, nous avons découragé les travailleurs qui étaient prêts à travailler à temps partiel préférant embaucher ceux qui étaient disponibles pour faire des semaines de 40 heures. Même si nous préférerions encore embaucher du personnel à plein temps, nous réalisons que beaucoup de candidats compétents ne pourront travailler qu'à temps partiel. Dans nos prochains recensements d'essai, nous évaluerons différents taux de rémunération que nous pourrions offrir aux recenseurs, par exemple, une rémunération horaire ou à la pièce, des primes au rendement et par remboursement des coûts. Outre ces questions monétaires, nous savons que nous devons définir d'autres moyens d'accroître la motivation des employés. À cette fin, nous allons étudier la possibilité de mettre en oeuvre un programme d'enrichissement des tâches afin de varier le travail des employés temporaires. Nous allons également essayer de mettre au point de nouvelles stratégies pour recruter des employés compétents et motivés en faisant appel à divers organismes communautaires, à des organismes à but non lucratif, à des groupes sociaux et à des groupes volontaires.

Programme d'information et de promotion

Sans la collaboration du public, il serait impossible d'effectuer un recensement valable, dans un délai et à un coût raisonnables. Dans les recensements antérieurs, nous avons obtenu une excellente collaboration du public en insistant sur l'importance du recensement dans nos campagnes de publicité et sur la stricte confidentialité des données de chaque répondant. Une campagne de promotion efficace permet d'accroître le nombre de questionnaires retournés par la poste, donc de réduire le nombre de suivis sur place, qui sont très coûteux. Pour le recensement de 1990, nous estimons que chaque point de pourcentage de plus dans le taux de réponse par la poste permettra de réaliser une économie de 5 à 6 millions de dollars.

L'élément le plus important de notre programme de promotion en 1980 a été le programme de publicité conçu par une agence choisie par l'Advertising Council. Ce programme a fourni pour 38 millions de dollars de temps d'antenne au Census Bureau et ce, sans aucun frais hormis les frais administratifs. Nous avons demandé à l'Advertising Council d'organiser notre prochaine campagne de promotion et ils ont accepté. Comme nous nous y sommes pris bien à l'avance, le conseil pourra participer à nos recensements d'essai et, de cette manière, mettre au point et tester différents concepts de promotion.

En plus de notre campagne de publicité, nous allons mettre en oeuvre différents programmes en collaboration avec les autorités locales. Dans notre Local Review Program, les autorités locales auront l'occasion de vérifier nos chiffres précensitaires de logements ainsi que nos chiffres postcensitaires de population et de logements et pourront nous signaler tout écart entre nos chiffres et les leurs. Le Census Bureau enverra ensuite du personnel sur le terrain pour contrôler ces écarts et, au besoin, nous effectuerons des corrections.

En outre, nous demanderons encore une fois aux communautés de mettre sur pied des Complete Count Committees. Au recensement de 1980, plus de 4,000 comités semblables ont été créés pour accroître la participation du public grâce à la collaboration des ressources locales. Nous estimons qu'ils ont été très utiles et nous voulons améliorer ce programme pour le recensement de 1990. Dans le recensement d'essai que nous avons effectué à Tampa en 1985, c'est en partie grâce aux efforts du Complete Count Committee mis sur pied par le maire de la ville qu'on a réussi à convaincre diverses vedettes locales (par exemple, des entraîneurs et des membres d'équipes sportives de Tampa) de prendre part à des messages publicitaires sur le recensement. Le maire lui-même à participé à un message publicitaire à la télévision.

Enfin, nous allons organiser un certain nombre d'autres activités de sensibilisation. Nos bureaux régionaux ont des contacts réguliers avec de nombreux organismes communautaires et ils en accroîtront la fréquence au fur et à mesure qu'approchera le recensement de 1990. Nous mettrons également sur pied un programme d'information dans les établissements d'enseignement. Avec la collaboration des administrateurs scolaires, nous allons sensibiliser les élèves à l'importance du recensement et leur montrer comment répondre au questionnaire. En juillet 1985, nous avons rencontré des enseignants venus de toutes les régions du pays pour discuter avec eux de la forme que pourrait prendre le programme en 1990. Enfin, nous allons demander à des organismes religieux d'encourager leurs membres à participer au recensement.

Contenu du questionnaire

Je voudrais, en dernier lieu, vous parler du contenu du questionnaire. Comme le recensement doit permettre de répondre aux besoins en données des utilisateurs pendant au moins dix ans, la conception du questionnaire constitue un des volets les plus importants de la planification.

Dans nos consultations avec les utilisateurs, nous avons constaté que la diversité des données demandées était beaucoup trop grande pour ce que nous pouvons raisonnablement espérer satisfaire. Ces demandes sont légitimes pour la plupart et portent sur une vaste gamme de sujets couvrant tous les aspects de notre société complexe. Par exemple, nous avons reçu dernièrement des douzaines de lettres de tous les coins du pays nous demandant d'inclure des questions sur les animaux de compagnie. Les auteurs de ces lettres nous ont présenté des arguments très valables pour montrer l'utilité de telles données dans la planification des programmes de santé publique et des programmes de stérilisation des animaux de compagnie.

Toutefois, un de nos objectifs pour le recensement de 1990 est de trouver un juste milieu entre les besoins des utilisateurs et le fardeau de réponse. Il est indispensable de trouver un tel compromis car la réussite du recensement dépend de la collaboration du public. Cette collaboration risque d'être compromise si le questionnaire est trop long. En pratique, cela signifie que nous ne pouvons pas beaucoup allonger le questionnaire en 1990.

Afin de décider quels sujets seront inclus dans le questionnaire, nous avons défini les sept critères suivants:

- Premièrement, nous n'allons recueillir que les données essentielles, c'est-à-dire les données dont la collecte est exigée par la constitution ou par la loi, les données particulières qui servent à l'application de programmes du gouvernement fédéral, des gouvernements des États et des administrations locales ainsi que les données qui permettront le mieux de décrire les principaux aspects de la situation du logement et de l'état de la population aux États-Unis.

- Deuxièmement, le recensement doit permettre de répondre aux besoins en données sur les petites régions. Le recensement convient bien à la production de données sur de petites régions, par exemple, sur le secteur de recensement dont la taille moyenne est de 4,000 habitants. Par contre, pour produire des données sur des régions plus importantes, par exemple, sur l'ensemble du pays, les États ou les grandes régions métropolitaines statistiques, les enquêtes-échantillon pourraient mieux convenir.

- Troisièmement, nous allons étudier la possibilité de rassembler des données sur des sous-populations peu nombreuses éparpillées sur le territoire. Le recensement convient mieux pour la collecte de telles données parce qu'une enquête-échantillon nationale ne donnerait pas une couverture convenable de ces groupes et, par conséquent, ne permettrait pas de recueillir des données statistiques fiables.

- Quatrièmement, comme nous allons privilégier la méthode d'autodénombrement, la majorité des recensés rempliront eux-mêmes le questionnaire, sans l'aide d'un recenseur. Les questions doivent donc être faciles à comprendre et formulées avec des mots que tout le monde connaît.

- Cinquièmement, les questions ne doivent pas poser de difficultés insurmontables du point de vue du traitement informatique, et les réponses doivent être relativement faciles à traduire sous une forme assimilable par machine. Cela signifie, entre autres choses, que le nombre de questions où l'on demandera au répondant de donner une réponse par écrit doit être réduit au minimum.

- Sixièmement, nous ne retiendrons aucune question qui, selon nous, peut indisposer le recensé ou envahir sa vie privée ou encore provoquer des controverses. Le Census Bureau a besoin de la collaboration du public et ne prendra pas le risque de perdre cette collaboration en posant des questions inappropriées.

- Septièmement, dans le questionnaire de 1990, nous reprendrons beaucoup de questions qui figurent dans le questionnaire de 1980 et dans les recensements précédents. Les réponses aux questions en 1990 pourront nous fournir les données dont nous avons besoin pour analyser l'évolution des principales caractéristiques socio-économiques et du logement. Cela ne veut pas dire qu'au recensement de 1990 nous ne supprimerons aucune question posée en 1980 ni que nous n'en inclurons pas de nouvelles. Nous veillerons cependant à maintenir la continuité et la comparabilité des données recueillies lors de chacun des recensements précédents.

Ce mois-ci, nous venons de terminer une série de rencontres au niveau local. Dans ces réunions, des utilisateurs de tous les coins du pays nous ont fait leurs suggestions sur le contenu du questionnaire ainsi que sur les produits et les régions géographiques de recensement. Dans chaque État, nous avons eu au moins une réunion avec les utilisateurs. Nous avons également terminé les rencontres que nous avions prévues avec les représentants des organismes fédéraux afin de déterminer les données dont ces derniers ont besoin pour administrer les programmes fédéraux. Nous sommes en train d'analyser les résultats de ces rencontres.

Dans les deux ou trois prochaines années, nous aurons de nombreuses décisions à prendre en ce qui concerne le contenu du questionnaire. À l'heure actuelle, nous préparons un essai national sur le contenu du questionnaire qui aura lieu en 1986 et qui permettra de vérifier les nouvelles questions et leur formulation. La loi nous oblige à soumettre au Congrès, au plus tard le 1er avril 1987, un rapport sur la nature des questions que nous souhaitons recueillir dans le recensement de 1990 ainsi que les questions proprement dites qui seront posées le 1er avril 1988.

Conclusion

Même s'il n'y a pas de changement à prévoir ni de modifications nécessaires à apporter au prochain recensement, la tenue d'un recensement constitue un défi énorme et une expérience stimulante. En 1980, nous avons dénombré plus de 226 millions d'habitants et 88 millions de logements et rassemblé des renseignements détaillés sur les caractéristiques de ces habitants et de ces logements. La tâche n'a pas été facile, en partie à cause de la très grande mobilité de la population aux États-Unis et de la diversité des conditions de vie. La tâche sera d'autant plus difficile en 1990 puisque, d'après les estimations, il y aura aux États-Unis 24 millions d'habitants et 18 millions de logements de plus à dénombrer.

En outre, nous devons non seulement dénombrer les personnes qui vivent dans des maisons, des appartements, des condominiums, des maisons mobiles, etc., mais aussi les personnes qui habitent des logements collectifs comme des casernements militaires, des résidences universitaires, des établissements pénitentiaires, des hôpitaux pour malades chroniques et des camps de travailleurs agricoles migrants. Nous devons même dénombrer les personnes qui n'ont pas de logis.

Dans ce document, nous avons examiné les grandes questions auxquelles le U.S. Bureau of the Census doit répondre pour préparer le recensement de la population et du logement de 1990. Bien que ces cinq questions: procédures de base, automatisation, recrutement du personnel, campagne d'information et de promotion et contenu du questionnaire, soient parmi les plus importantes, il reste des milliers d'autres questions et de réponses qui rendent le recensement complexe et stimulant.

Nous désirons discuter de ces questions et échanger sur les problèmes et les solutions avec nos collègues d'autres pays. Nous sommes également désireux de connaître les principaux problèmes qu'ils doivent résoudre dans la préparation de leurs prochains recensements.

PLANIFICATION DES PROCHAINS RECENSEMENTS DE LA POPULATION EN AUSTRALIE

HENRY KRIEGEL

SYSTÈMES DE RECENSEMENT DE LA POPULATION
ÉLABORATION ET PLANIFICATION À LONG TERME
BUREAU DU RECENSEMENT DE L'AUSTRALIE

Au cours des deux prochaines années, l'Australian Bureau of Statistics (ABS) se propose d'examiner un certain nombre de questions touchant les modalités des recensements qui auront lieu après 1986. Ces questions sont très diverses; elles peuvent être conceptuelles ou méthodologiques ou avoir trait à certains aspects du contenu du questionnaire, qu'il faudra, bien entendu, étudier de manière plus approfondie à une étape ultérieure. L'objectif de ce document est de discuter des principales questions que l'ABS se propose d'examiner.

Selection des sujets

Pour les recensements de 1981 et de 1986, un des aspects importants de la sélection et du développement des questions à inclure dans le recensement a été la création d'un processus de consultation. Ce processus a permis d'inciter le public et les principaux utilisateurs à soumettre des suggestions sur des sujets d'intérêt.

Après examen des suggestions, l'ABS a effectué des essais pour les sujets les plus importants. Il a ensuite dressé la liste des sujets qu'il jugeait nécessaire d'inclure dans le recensement et, après évaluation des réactions des principaux utilisateurs, a annoncé ses recommandations finales. L'ABS a également retenu, dans le recensement de 1986, des suggestions faites pour celui de 1981.

Les utilisateurs consultés pour le recensement de 1986 n'ont proposé l'inclusion d'aucun nouveau sujet présentant beaucoup d'intérêt. En outre, les suggestions reçues contenaient très peu d'éléments d'information nouveaux sur les besoins des utilisateurs et sur les raisons pour lesquelles ces derniers demandent l'inclusion de certains sujets. Pour le prochain recensement, il est peu probable que l'on en apprenne davantage en demandant aux utilisateurs de faire des suggestions. Il faut donc trouver d'autres moyens d'évaluer les besoins nouveaux sans toutefois supprimer le procesus de consultation, qui est important. Il serait plus utile de soumettre aux utilisateurs les sujets qui pourraient être inclus dans le recensement et de leur demander leur avis, puis, compte tenu des tendances observées depuis un certain nombre d'années, de déterminer s'il convient mieux de recueillir les diverses données au moyen d'un recensement quinquennal ou au moyen d'enquêtes. Il faudrait également envisager la possibilité d'utiliser cette évaluation dans la planification des enquêtes-ménages.

Pour faire notre choix, nous pourrions diffuser simultanément un ensemble de documents de travail portant chacun sur un sujet susceptible d'être inclus dans le recensement. Ces documents ressembleraient à nos rapports préliminaires. Toutefois, pour étayer nos arguments, nous y effectuerions une analyse plus approfondie des tendances observées depuis un certain nombre d'années. En outre, dans chaque document, nous examinerions de plus près la question de savoir ce qui constitue le meilleur moyen de répondre aux besoins des utilisateurs: le recensement ou les enquêtes. Les commentaires des utilisateurs et du public concernant les documents de travail, et en particulier en ce qui a trait à l'inclusion de nouveaux sujets, serviraient à l'élaboration des recommandations finales.

Dénombrement

Dans tous les recensements tenus jusqu'à maintenant en Australie, les personnes ont été dénombrées à l'endroit où elles se trouvaient le soir du recensement. Jusqu'en 1971, toutes les statistiques du recensement relatives à une région donnée représentaient les personnes qui s'y trouvaient ce soir-là. C'est au recensement de 1976 que nous avons inclus pour la première fois une question sur le lieu habituel de résidence afin d'être en mesure d'évaluer la migration interne. Les données recueillies ont servi à produire un nombre très limité de tableaux sur le lieu habituel de résidence. La question a été reprise dans le recensement de 1981 afin d'établir une estimation de la population résidante et de produire un ensemble limité de tableaux sur le lieu habituel de résidence.

Aux recensements de 1976 et de 1981, on a attribué aux adresses de lieu habituel de résidence un code correspondant au secteur de recensement local (census local government area - LGA). En 1986, nous changerons d'unité et utiliserons le secteur statistique local (statistical local area - SLA). On peut, par ce moyen, produire

des tableaux sur le lieu habituel de résidence pour des régions formées par regroupement de LGA. Pour le recensement de 1986, nous les produirons par regroupement des SLA. En 1981, nous avons commencé à produire une série d'estimations sur la population résidante et, depuis, il y a eu accroissement de la demande de données établies à partir du lieu habituel de résidence. Au recensement de 1986, nous serons plus en mesure de répondre à cette demande accrue en produisant une gamme étendue de données établies en fonction du lieu habituel de résidence.

Toutefois, compte tenu des méthodes actuelles, il y a des limites à la quantité de données qui peuvent être produites à partir du lieu habituel de résidence. Nous ne pouvons pas produire ces données pour les régions dont les limites ne correspondent pas à celles d'un SLA ou d'un groupe de SLA ni de données sur l'ensemble des caractéristiques des familles et des ménages dont un ou des membres étaient absents temporairement le soir du recensement. Pour éliminer certaines lacunes dans les statistiques de la famille, nous avons inclus dans le questionnaire de 1986 une question sur les résidents habituels absents temporairement le soir du recensement. Seules les données démographiques de base de ces personnes seront recueillies en vue d'améliorer le dénombrement des familles par type. Cependant, cette question ne permettra pas de corriger d'autres lacunes dans les statistiques de la famille, notamment celles observées relativement au revenu.

Des données établies en fonction du lieu habituel de résidence pourraient être produites pour plus de régions en attribuant à l'adresse habituelle de résidence du recensé un code correspondant au district de collecte, mais cela coûterait cher. Par ailleurs, pour corriger les lacunes dans les données sur les familles et les ménages, il faudrait attribuer au lieu habituel de résidence les données recueillies sur les personnes et les familles absentes de leur résidence habituelle le soir du recensement mais dénombrées ailleurs. L'opération pourrait être effectuée manuellement au centre de transcription des données mais elle retarderait probablement le traitement et la diffusion des premières données et serait très coûteuse. On pourrait également effectuer cette opération par ordinateur en mettant provisoirement en mémoire les adresses des personnes absentes. Cela ne coûterait que l'opération d'introduction, mais il faudrait changer la politique actuelle, qui est de ne jamais mettre d'adresses en mémoire. Or ce changement est peu probable.

Si nous dénombrions les personnes à leur résidence habituelle, nous éviterions bon nombre des problèmes posés par la méthode de dénombrement employée actuellement dans les recensements australiens. Le dénombrement au lieu habituel de résidence est la méthode employée aux États-Unis, au Canada, au Japon et dans la plupart des pays d'Europe. Si nous changions notre méthode, toutes les données relatives aux membres d'un ménage figureraient sur un seul questionnaire, ce qui éliminerait une partie des frais de traitement. Il y aurait toutefois rupture dans les séries chronologiques de données utilisées pour les études longitudinales et il se pourrait que l'on perde les données sur les résidents séjournant dans une région quelconque le soir du recensement (ces données sont très importantes pour les utilisateurs qui s'intéressent à des régions où le nombre de personnes de passage est toujours élevé, par exemple les centres de villégiature). De plus, ce changement de méthode de dénombrement pourrait créer une certaine confusion chez les recensés.

Même si nous pouvons profiter de l'expérience d'autres pays, un changement de méthode de dénombrement entraînerait des coûts de développement très élevés. Parmi les problèmes qu'il faudrait résoudre figure le dénombrement des Australiens séjournant à l'étranger le jour du recensement ou ceux n'ayant pas de lieu habituel de résidence.

Géographie statistique du recensement

Les divers aspects de la géographie statistique du recensement qui touchent la production des données du recensement devront être examinés afin de déterminer si des changements ou des améliorations s'imposent.

a) La demande accrue de données du recensement sur les petites régions s'est traduite par une plus grande utilisation du district de collecte comme plus petite unité pour laquelle on puisse produire des données. Les districts de collecte sont également les unités employées pour produire des données sur les régions définies par les utilisateurs. Toutefois, les districts de collecte ont été créés essentiellement dans le but de mieux gérer le travail effectué sur le terrain et leur utilisation pose des difficultés au niveau de la diffusion des données. D'une part, les districts de collecte sont généralement trop grands pour que l'on puisse les regrouper et obtenir des données exactes sur les régions géocodées plus importantes et, d'autre part, compte tenu des contraintes de confidentialité, ils sont trop petits pour permettre la diffusion des données détaillées sur les caractéristiques de base.

Pour produire des données sur toutes les petites régions auxquelles les utilisateurs sont susceptibles de s'intéresser, et dont l'étendue varie considérablement, il faudrait, au minimum, attribuer un code d'îlot aux adresses des ménages et, idéalement, un géocode. Si l'on décide d'attribuer des codes d'îlot, il faudra établir, pour l'ensemble du pays, des cartes d'îlots dont les limites coïncident avec celles des secteurs de collecte. Ces cartes pourraient être produites sans trop de frais. Toutefois, en raison des contraintes de confidentialité, la quantité de données que nous pourrions fournir aux utilisateurs serait bien inférieure à celle que nous pouvons diffuser en ce moment à partir des districts de collecte. La mise en mémoire des données à partir de cartes d'îlots faciliterait l'extraction et, de cette façon, l'ABS serait mieux en mesure d'agréger les données pour fournir aux utilisateurs des statistiques plus justes sur les régions définies par ces derniers. Dans cette hypothèse, l'ABS serait toutefois obligé de consacrer des ressources beaucoup plus considérables pour répondre aux demandes ad hoc des utilisateurs alors que depuis quelques recensements il les encourage à se procurer des fichiers sommaires pour agréger leurs données. Il faudrait également prendre des mesures particulières pour s'assurer que les utilisateurs ne puissent obtenir de données leur permettant d'identifier certains recensés lorsqu'ils retranchent certaines des données recueillies pour une petite région afin de produire des statistiques pour une région encore plus petite.

On pourrait effectuer le géocodage des adresses en demandant aux recenseurs d'indiquer sur leurs cartes l'emplacement du logement de chaque ménage. On attribuerait ensuite à chaque logement un code numérique établi à partir du code du district de collecte. Toutefois, comme le géocodage implique la mise en mémoire des adresses, il faudrait modifier notre politique sur cette question.

Afin de diffuser des données plus détaillées, on pourrait également envisager de créer une nouvelle unité statistique qui serait plus grande que le district de collecte mais dont la taille ne dépasserait pas 10,000 habitants. La taille et l'utilisation de cette unité correspondraient à peu près à celles des secteurs de recensement américains. Dans notre classification géographique type (Australian Standard Geographical Classification – ASGC)[1], cette nouvelle unité se situerait entre le district de collecte et la subdivision statistique. Les unités seraient donc vraisemblablement constituées par regroupement de petits SLA ou par fractionnement de gros SLA.

b) Les critères relatifs à l'établissement des secteurs normalisés ont été définis il y a 20 ans. Depuis leur adoption, ils ont fait l'objet d'un certain nombre de changements mineurs, en particulier pour permettre la création des districts statistiques. Nous considérons que ces critères doivent être évalués en fonction de la répartition de la population et des besoins des utilisateurs.

c) La création de zones de destination normalisées pour les données sur le lieu de travail qui seraient compatibles avec l'ASGC permettrait de mettre en mémoire les codes de destination des unités figurant dans le registre intégré des entreprises de l'ABS. La création de ces zones permettrait peut-être de réduire les dépenses qu'il faut engager à chaque recensement au titre du matériel de codage et de produire des estimations intercensitaires de l'emploi selon la zone de destination (après avoir tenu compte du sous-dénombrement dans le registre).

d) Si, en guise d'adresse, nous utilisions le code postal (circonscriptions postales) et le nom de la banlieue indiqués par le recensé sur son questionnaire, nous pourrions assurer une compilation plus précise des données sur de tels secteurs. À l'heure actuelle, pour produire des données sur ces secteurs, nous faisons des approximations à partir de districts de collecte complets. Les circonscriptions postales et les banlieues ne sont pas incluses dans l'ASGC, sauf dans les cas où les limites des banlieues correspondent à celles des SLA. Cependant, la demande de données du recensement sur des secteurs de ce genre continue d'augmenter au fur et à mesure qu'augmente la quantité de données produites à leur sujet par des organismes administratifs. Soulignons qu'on pourrait également répondre à cette demande de données grâce aux points de référence géographique advenant leur implantation.

En plus de s'occuper de ces questions qui sont particulières au recensement, la section responsable de la classification géographique à l'ABS doit spécifier, relier et rationaliser dans la mesure du possible, les autres régions et uniformiser toutes les classifications dont

1 Note du rédacteur.

nous nous servons pour produire des statistiques (par exemple, les paroisses dans l'État de Victoria, et les circonscriptions postales) mais qui, pour des raisons techniques ou autres n'ont pas été incluses dans l'ASGC. Les résultats de ce travail paraîtront dans un manuel de référence sur la géographie statistique qui s'intitulera "Geographical Classification Framework".

Cartographie

La Division of National Mapping, qui prépare les cartes pour le travail sur le terrain ainsi que celles qui serviront à la diffusion des données pour chaque recensement, prévoit mettre sur pied une importante base de données géographiques qui comprendra des données sur cartes numérisées et de mettre au point un système de cartographie assisté par ordinateur. L'adoption de ces mesures devrait nous permettre de réduire le coût de la préparation des cartes pour chaque recensement et nous offre un moyen rapide et souple de produire toutes les cartes nécessaires pour le travail sur le terrain et pour la diffusion des données. L'ABS a été invité à participer à ces développements.

La création d'une base de données portant sur l'ensemble du territoire australien en faisant appel aux systèmes d'information géographique des administrations des États et des administrations locales permettrait de maintenir les coûts de développement à un niveau raisonnable. Cette base nous permettrait de dresser une liste de tous les logements, de leur attribuer un code unique et pourrait, à long terme, permettre d'effectuer des recensements basés sur la méthode de retour par la poste, d'apparier les personnes et les logements à leur lieu habituel de résidence et aussi de produire plus aisément des données sur des régions définies par les utilisateurs. Ces questions sont soulevées dans d'autres sections de ce document.

Recensements utilisant la méthode de retour par la poste

Avant de commencer la préparation du recensement de 1986, nous avons fait une étude de la méthode de retour par la poste, principalement pour déterminer si cette dernière n'assurerait pas une meilleure protection des formules de recensement. L'étude a montré que les recensements de retour par la poste pourraient être moins coûteux que la méthode actuelle de livraison et de reprise si plus de 80% des ménages retournaient leur questionnaire dans les deux semaines suivant le jour du recensement. Les recenseurs effectueraient un suivi auprès des ménages n'ayant pas rempli le questionnaire ou ayant retourné un questionnaire trop incomplet. L'étude a également montré que moins les cas de non-réponse et de réponse trop incomplète seraient nombreux, plus la méthode de retour par la poste serait économique.

Dans cette étude, nous n'avons pas examiné la question de savoir si la méthode de retour par la poste, en nous déchargeant de la livraison et du gros des opérations de reprise et en nous assurant une économie de temps appréciable, ne nous permettrait pas d'accroître la taille des districts de collecte. Avec la méthode de retour par la poste, le dénombrement d'un district de collecte de plus grosse taille nécessiterait un personnel moins nombreux et, par conséquent, les opérations seraient moins coûteuses. En pratique cependant, la taille des districts de collecte ne serait vraisemblablement augmentée que dans les régions où il y aurait croissance de la population et cela serait fait en augmentant la taille maximale que peut atteindre un district avant d'être fractionné. Cet état de choses tient au fait qu'il faut maintenir la comparabilité des données par rapport aux recensements antérieurs et qu'il serait très onéreux de procéder à l'établissement de districts complètement nouveaux.

Suivi téléphonique

Aux États-Unis et au Canada, on effectue un suivi téléphonique auprès des ménages dont le questionnaire est trop incomplet afin d'améliorer la qualité des données et de réaliser la plus grande économie d'argent possible. Les suivis se font généralement à partir des bureaux régionaux qui contrôlent les opérations sur le terrain et, à l'occasion, à partir du bureau central de traitement des données.

Aux États-Unis, on prévoit recourir davantage au suivi téléphonique dans le recensement de 1990 en mettant au point un système automatisé d'enregistrement des questionnaires de retour par la poste. Ce système impliquerait un accroissement de la charge de travail des opérateurs effectuant des interviews téléphoniques assistées par ordinateur (CATI). En Australie, ce sont les recenseurs qui effectuent les suivis lorsqu'ils vont reprendre les questionnaires. Si nous adoptons la méthode de retour par la poste pour nos recensements, nous devrons décider si nous allons demander aux recensés d'indiquer leur numéro de téléphone sur le questionnaire pour nous permettre d'effectuer des suivis téléphoniques et, de cette façon, de supprimer les suivis sur le terrain qui sont coûteux. Il faudra donc faire l'essai de la méthode de retour par la poste afin de déterminer si les personnes sont disposées à indiquer leur numéro

de téléphone et à fournir des renseignements au téléphone. Il faudra également évaluer la fréquence des cas où il serait impossible de joindre les recensés par téléphone.

Traitement des données

Nous avons commencé à étudier d'autres méthodes de saisie des données pour les prochains recensements. Mis à part les questions reliées à la nature même du travail et aux problèmes répétitifs dont la cause est le stress, l'introduction manuelle des données, méthode que nous allons utiliser pour le recensement de 1986, exige une main-d'oeuvre nombreuse et coûte très cher. Il serait peut-être avantageux d'utiliser des marques optiques, c'est-à-dire d'inclure dans le questionnaire le plus grand nombre possible de questions dont les réponses inscrites par le recensé pourraient être décodées par un lecteur optique de marques. Quant aux autres questions, c'est-à-dire à celles pour lesquelles le recensé doit fournir une réponse en toutes lettres, nous pourrions inscrire les codes et les marques appropriés au bureau de traitement avant que le questionnaire soit décodé par le lecteur.

À défaut de cette méthode, nous pourrions peut-être envisager la solution suivante: si, au moins pour les questions touchant les données provisoires servant à produire nos estimations démographiques, nous pouvions prévoir des réponses décodables par lecteur optique de marques, nous pourrions peut-être éliminer les coûts très élevés du traitement préliminaire des données qui sert à faire les révisions voulues aux estimations démographiques. Toutes les réponses qui seraient elles-mêmes un code seraient décodées par un lecteur optique de marques; cela permettrait de constituer un fichier d'enregistrements partiels (partiels en ce que la saisie des réponses en toutes lettres n'y serait pas effectuée). Grâce à ces enregistrements partiels, nous pourrions produire les chiffres provisoires du recensement nécessaires à l'établissement des estimations démographiques. Par la suite, des commis pourraient reprendre les fichiers d'enregistrements partiels et effectuer, en mode interactif, la saisie, le codage et (ou) le contrôle des renseignements fournis en toutes lettres qu'ils verraient sur le questionnaire même ou sur un autre support, par exemple, un microfilm ou encore un enregistrement vidéo produit au moment de la lecture des marques. La France a déjà fait l'essai d'un système semblable tandis que la Nouvelle-Zélande entend faire de même en 1986 et les États-Unis en 1990.

Nous allons examiner de plus près la possibilité d'utiliser le codage informatisé et (ou) le codage assisté par ordinateur en mode interactif. En plus de nous aider à assurer l'uniformité du codage, ces deux systèmes nous permettraient d'alléger l'activité intensive que requiert la tâche du codage manuel des données. Ces systèmes pourraient faire partie d'une approche méthodologique globale telle qu'elle semble être perçue par la France, la Nouvelle-Zélande et les États-Unis.

Une autre question qui se pose au sujet du traitement des données est celle de la centralisation par opposition à la décentralisation des opérations. Le centre de transcription des données établi pour les derniers recensements et pour celui de 1986 centralisait les opérations et a permis de réaliser des économies considérables et de produire des données de qualité uniforme. Toutefois, comme les facteurs dont il faut tenir compte pour établir le coût des opérations peuvent beaucoup changer avec le temps (par exemple, en ce qui concerne le matériel, les réseaux de communication, les besoins en personnel lorsque les opérations sont centralisées), il serait peut-être bon d'envisager une certaine décentralisation du traitement des données et des opérations manuelles. Les problèmes d'uniformité du codage qu'ont connus plusieurs pays qui ont décentralisé le traitement des données pourraient dans une large mesure être évités au moyen du codage automatique ou assisté par ordinateur.

Résumé de la discussion

La seconde partie de la session sur le rôle de la recherche et des essais dans le cycle du recensement a consisté en une table ronde ayant pour thème "Maintenant, que faisons-nous?" Avant de commencer la discussion, les panelistes ont répondu à quelques questions de l'assistance.

1. Questions de l'assistance

Une des questions portait sur les difficultés qu'éprouvent certains pays à obtenir les noms et adresses des recensés et sur les problèmes posés par la conservation des documents du recensement. Les panelistes de tous les pays ont affirmé qu'il était indispensable, pour des raisons opérationnelles, d'obtenir les noms et adresses des recensés mais ils ont ajouté que la période pendant laquelle ces renseignements sont conservés varie largement d'un pays à l'autre. L'Australie, par exemple, détruit tous ses documents dans les deux années qui suivent le recensement. Le Canada et le Royaume-Uni conservent leurs documents de recensement plusieurs années, ceci principalement à des fins de recherche historique. Cependant les deux pays ont pour politique de ne pas inscrire les noms et adresses des recensés dans les fichiers informatisés. Enfin, on souligne que le public est loin d'être prêt à accepter les moyens d'emmagasinage de l'information offerts par la nouvelle technologie, par exemple, la formation de registres d'adresses ou le couplage de fichiers.

Une autre question portait sur l'indépendance des instruments qu'il conviendrait d'utiliser pour mesurer la couverture et pour évaluer le contenu du recensement. Comme les variables de base (par exemple, l'âge et le sexe) sont souvent nécessaires pour l'appariement et aussi parce qu'elles peuvent contenir des erreurs, l'utilisation du même instrument pour évaluer l'une et l'autre peut fausser la valeur de cette mesure. On a par ailleurs souligné qu'une évaluation de la couverture au moyen d'une enquête postcensitaire peut produire une sous-estimation de la population si les mêmes ménages sont oubliés. Une autre question adressée au panel portait sur l'utilisation des enquêtes sur les ménages pour faire l'essai de questions susceptibles d'être retenues dans un recensement. L'Australie effectue des enquêtes mensuelles dans son processus de sélection des sujets. Toutefois, dans la plupart des pays, l'enquête n'est pas souvent utilisée à des fins d'essai en raison des différences dans les méthodes de collecte appliquées dans le cas d'un recensement et d'une enquête.

Enfin, les panelistes se sont livrés à un bref échange de vues sur le coût des essais par rapport à l'ensemble des coûts des opérations d'un recensement. Dans la plupart des pays, les essais représentent de 2 à 10% du coût total des opérations.

2. Table ronde: Maintenant, que faisons-nous?

Tous les panelistes s'entendent pour dire que les recherches et les essais constituent un élément important dans la planification d'un recensement. Toutefois, les difficultés auxquelles les différents pays font face dans la conception de tels programmes ne sont pas les mêmes puisque leur cycle de recensement diffère. **M. Bounpane** *prend la parole en premier lieu et explique la façon dont le U.S. Bureau of the Census a procédé pour mettre au point son programme d'essais.*

Aux États-Unis, on a procédé en trois étapes pour préparer le programme d'essais. Premièrement, on a défini les problèmes qui s'étaient posés au recensement de 1980 et auxquels il fallait trouver une solution. Deuxièmement, on a constitué un groupe de comités internes chargés d'étudier des sujets spécifiques et, pour chacun d'eux, d'évaluer les méthodes utilisées dans le passé et de définir les méthodes qui devraient être appliquées pour le prochain recensement. Chaque comité a déposé un rapport. Troisièmement, on a tâché d'obtenir des suggestions de l'extérieur sur le recensement.

Tout cela faisait une très longue liste de sujets susceptibles d'être retenus pour le recensement, qu'il a donc fallu réduire. Cette tâche a été accomplie, et pouvait uniquement l'être, par les spécialistes du Census Bureau. Les sujets pour lesquels on a décidé d'effectuer des essais sont ceux qui entraînaient le plus de changements et qui présentaient le plus de risques. N'ont pas fait l'objet d'essais les sujets que le Bureau jugeait pouvoir inclure dans le recensement sans trop de risques.

Le U.S. Bureau of the Census a tenté de créer un certain soutien à l'intérieur du Bureau. À cette fin, il a constitué un groupe de travail composé de spécialistes de différents domaines dont le rôle était d'évaluer les différents objectifs et les diverses options possibles pour le recensement ainsi que de formuler des recommandations sur les sujets à tester. Une des fonctions de ce groupe a permis la sensibilisation du personnel du Bureau aux différentes fonctions qu'exige la tenue d'un recensement.

***M. E.T. Pryor** a ensuite pris la parole pour donner un bref aperçu des composantes devant faire l'objet d'essais en vue du recensement de 1991 au Canada.*

Premièrement, le contenu est une des composantes du recensement pour laquelle il est indispensable de tenir des consultations et d'effectuer des essais. Le contenu du questionnaire influe sur tous les aspects du recensement et il faut le définir tôt dans le processus du recensement. Le contenu doit également faire l'objet des tout premiers essais.

Deuxièmement, l'automatisation des opérations doit également faire l'objet d'essais. Par exemple, il pourrait être avantageux d'automatiser la planification des opérations, la collecte, l'exploitation du système d'information de la gestion et le traitement des données.

Troisièmement, il faut créer un environnement qui favorise l'épanouissement des employés travaillant au recensement. Comment allons-nous recruter, former et motiver le personnel dont nous avons besoin, compte tenu des contraintes budgétaires et des limites imposées en matière de main-d'oeuvre?

Quatrièmement, la question de la confidentialité des données et du respect de la vie privée du recensé doit faire l'objet d'études en vue de l'élaboration d'un plan d'action.

Cinquièmement, il faut évaluer nos méthodes de collecte. Nous devons notamment nous pencher sur des questions comme les registres d'adresses, l'extension de la méthode de retour par la poste, la production des questionnaires dans les deux langues officielles, la centralisation des opérations de contrôle et l'échantillonnage.

Les mesures de qualité par rapport au sous-dénombrement ou au surdénombrement et au contenu ont, elles aussi, besoin de recherches et d'essais.

Enfin, nous devons examiner la question de la diffusion des données à la lumière d'une politique pour les produits du recensement, à la lumière aussi des nouvelles technologies et de la décentralisation de la fonction diffusion.

Pour atteindre nos objectifs, nous devons évidemment avoir un plan d'action très clair et ne pas perdre l'élan que nous avons pris à cette conférence.

3. Mot de la fin des panelistes

Pendant les dernières minutes de la conférence, les panelistes ont fait quelques dernières remarques et suggestions.

Le panel avait trois remarques générales à faire au sujet des recensements. Premièrement, pour l'avenir, il est très important dans la planification du recensement de 1991 de ne pas perdre de vue les recensements qui viendront par la suite. Deuxièmement, il faut se rappeler que ce qui constitue un problème aujourd'hui n'en sera peut-être pas un demain. Troisièmement, un recensement est une chose importante dans n'importe quel pays et c'est pourquoi il ne doit pas échouer. Les changements au recensement doivent donc être envisagés avec prudence.

En outre, les panelistes étaient tous d'accord pour dire que les pays qui font des recensements devraient multiplier et varier leurs échanges. Il devrait y avoir plus d'échanges de personnel; en particulier, il serait utile d'inviter des employés à venir travailler à un programme d'essais ou à un aspect particulier des opérations. Il faudrait également que chaque pays continue, mais de manière plus systématique, à échanger des renseignements et de la documentation avec les autres et à leur faire partager le résultat de ses expériences.

Il faudrait que chaque pays fasse plus souvent profiter les autres de ses expériences en matière d'essais. Comme il est difficile d'obtenir des fonds et de trouver le temps d'effectuer des essais, chaque pays devrait pouvoir tirer le maximum des expériences des autres.

ALLOCUTION DE FERMETURE

Edward T. Pryor

Directeur général de la Direction de la
statistique démographique et du recensement
Statistique Canada

J'aimerais profiter de l'occasion pour remercier sincèrement tous ceux qui ont participé à cette conférence.

En premier lieu, j'aimerais remercier nos invités de l'étranger d'avoir participé à la conférence, d'y avoir présenté des exposés et d'être intervenus dans les débats. Les points de vue qu'ils ont fait valoir nous ont permis de découvrir une foule de nouvelles idées et de nouvelles façons d'aborder les questions d'intérêt général et particulier.

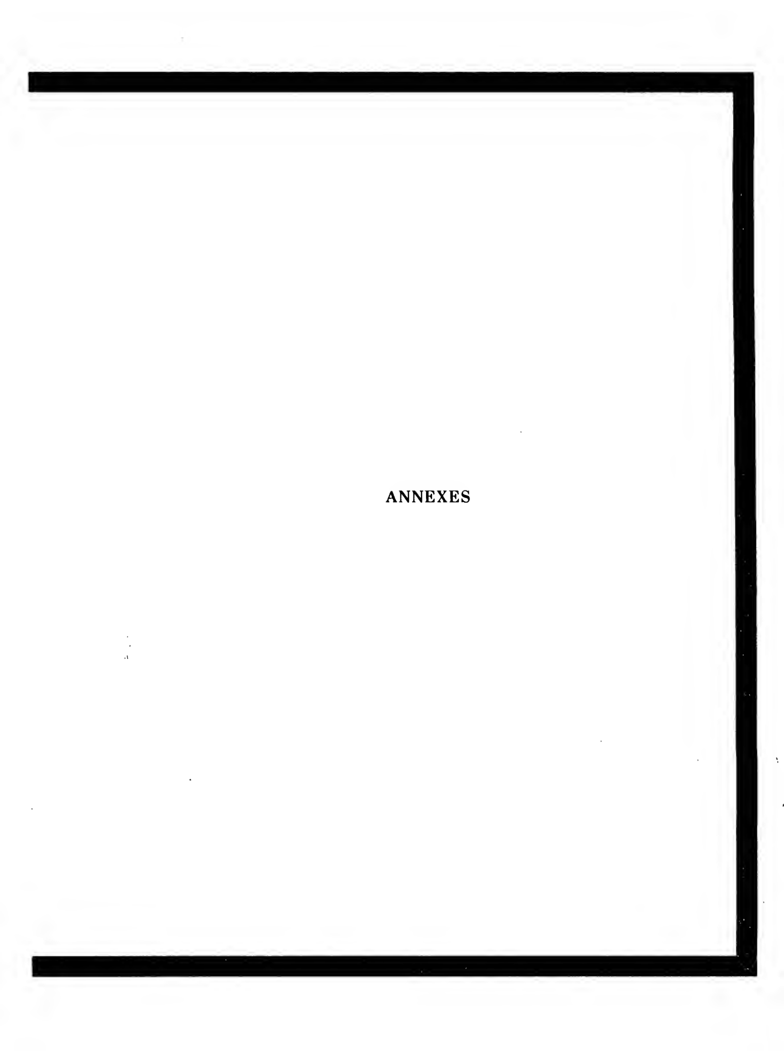
En deuxième lieu, j'aimerais remercier MM. Ivan Fellegi et Bruce Petrie de l'aide qu'ils m'ont apportée pour l'organisation de la conférence. Sans leur appui, la conférence ne se serait probablement pas concrétisée.

Bien sûr, mes remerciements s'adressent surtout à M. Don Royce, qui s'est chargé de prendre les arrangements nécessaires pour la tenue de la conférence. S'il ne s'était pas occupé des préparatifs, il aurait été presque impossible de tenir la conférence.

J'aimerais également remercier le comité du programme, composé de Ken Brown, Richard Ethier, Gilles Montigny, Henry Puderer, Doug Hicks et Wilson Freeman. Le programme qu'ils ont établi a grandement facilité le déroulement de la conférence.

Je tiens aussi à remercier le comité des arrangements locaux, composé de Kevin Thatcher, Karen Kelly, Kathy Miller, Owen Power, Anis Ashraf et Josée Dufresne. Nous leur sommes redevables d'une bonne partie du succès de la conférence.

Enfin, je désire remercier les participants de leur contribution enthousiaste à la conférence. Grâce à eux, Statistique Canada dispose d'ores et déjà d'une base de lancement du programme de 1991. Ce programme sera établi à partir des idées et des approches que vous n'avez pas hésité à proposer. Je vous en remercie.



ANNEXES

ANNEXE 1 - PROGRAMME

CONFÉRENCE SUR LA PLANIFICATION

DU RECENSEMENT DE 1991

DU 8 AU 11 OCTOBRE 1985

STATISTIQUE CANADA

Le mardi 8 octobre

8 h 00 - 9 h 00	Inscription et café	
9 h 00	Allocution de bienvenue:	I.P. Fellegi Statisticien en chef, Statistique Canada
9 h 15	Allocution d'ouverture:	E.T. Pryor Directeur général, Direction de la statistique démographique et du recensement
9 h 30 - 12 h 30	Séance: Collecte des données du recensement	
	Président:	John Riddle Directeur général associé, Opérations régionales
9 h 30	Introduction au processus de collecte du recensement – Présentation audio-visuelle du programme du recensement de 1986	
9 h 50	Automatisation de la collecte Brian Williams, directeur adjoint, Bureau régional du Manitoba et du Sud de la Saskatchewan, Opérations régionales Suivi téléphonique centralisé John Kazmaier, chef adjoint de la Division des recensements, Division des opérations régionales, Bureau du recensement des États-Unis	
10 h 45	Pause-café	
11 h 00	Extension du retour par la poste dans les secteurs de reprise Doug Hicks, Division des opérations des enquêtes	
11 h 20	Table ronde:	Les conditions relatives à la collecte de 1991: trois points de vue – Confidentialité/Point de vue légal Richard Barnabé, directeur régional, Opérations régionales du Québec – Groupes d'intérêt spécifique Boyd Underhay, directeur régional, Opérations régionales de Terre-Neuve et du Labrador

- Population autochtone et régions éloignées
  Jerry Page, directeur régional, Opérations régionales de l'Alberta, du Nord de la Saskatchewan et des T.N.-O.

14 h 00 - 17 h 00	Séance: <u>Géographie du recensement</u>	
	Président:	D. Ross Bradley Directeur, Division de la géographie
14 h 00	Soutien géographique pour le recensement décennal de 1990 Silla Tomasi, chef adjoint de division pour les opérations, Division de la géographie, Bureau du recensement des États-Unis	
14 h 50	Pause-café	
15 h 10	Fichier principal des régions: un meilleur moyen pour répondre aux besoins du recensement Joel Yan, Sous-division de la géocartographie Un programme d'îlots: oui ou non Robert Parenteau, Division de la géographie Les systèmes informatiques au service du géocodage du recensement Gordon Deecker, Sous-division de la géocartographie	
16 h 15	Participant:	S. Witiuk, directeur adjoint, Division des services et du développement informatiques
16 h 25	Discussion avec participation de l'assemblée	

Le mercredi 9 octobre

9 h 00 - 12 h 00	Séance: <u>Le contenu du recensement de 1991</u>	
	Président:	Ian Macredie Directeur, Division de l'analyse des enquêtes des ménages et du travail
9 h 00	Discussion:	Le processus de développement du contenu de 1991
	Participants:	Henry Kriegel, directeur, Planification à long terme et développement des systèmes du recensement de la population, Bureau de la statistique d'Australie Susan Miskura, chef, Division de la planification décennale, Bureau du recensement des États-Unis David Pearce, chef, Division du recensement, Bureau des recensements de la population et des enquêtes démographiques, Royaume-Uni
10 h 00	Pause-café	

10 h 15	Table ronde:	Propositions sur le contenu de 1991
	Orateurs:	John Kettle, Futuresearch Publishing Inc.
		Frank Clayton, Clayton Research Associates
		Noah Meltz, Université de Toronto
11 h 45	Discussion avec participation de l'assemblée	
13 h 30 - 16 h 30	Séance: **Le recensement de l'agriculture**	
	Président:	Terry Gigantes Directeur général, Direction de la statistique sur les ressources, la technologie et les services
13 h 30	Dénombrement par la poste du recensement de l'agriculture aux États-Unis Charles Pautler, chef, Division de l'agriculture, Bureau du recensement des États-Unis	
	Pour un recensement de l'agriculture fondé sur l'emplacement de l'exploitation Oliver Code, Division de l'agriculture et des ressources naturelles	
14 h 30	Pause-café	
15 h 00	Méthodes de protection de la confidentialité des données du recensement de l'agriculture de 1991 Rick Burroughs, Division de l'agriculture et des ressources naturelles Mary March, Division des méthodes de recensement et d'enquêtes-ménages	
15 h 30	Participant:	Mel Jones, chargé du recensement, recensement de l'agriculture de 1986
16 h 00	Discussion avec participation de l'assemblée	

Le jeudi 10 octobre

9 h 00 - 12 h 00	Séance: **Automatisation du recensement**	
	Président:	Martin Podehl Directeur, Division des services et du développement informatiques
9 h 00	Plans d'automatisation pour le recensement de la population et du logement de 1990 aux États-Unis Peter A. Bounpane, directeur adjoint des recensements démographiques, Bureau du recensement des États-Unis	
	Méthodes possibles de saisie des données Dave Croot, Division des services aux clients	

La décentralisation des méthodes de la saisie des données pour le recensement des États-Unis
Arnold Jackson, chef, Division des opérations décennales, Bureau du recensement des États-Unis

10 h 30 Pause-café

10 h 45 Codage automatisé au Bureau de la statistique de la Suède
Lars Lyberg, Bureau de la statistique de la Suède

Logiciels généralisés
T. Mike Jeays, Division des services et du développement informatiques

11 h 30 Participant: J. Ryten, statisticien en chef adjoint, Secteur de l'informatique et de la méthodologie, Statistique Canada

11 h 45 Discussion avec participation de l'assemblée

13 h 30 - 16 h 30 Séance: <u>La couverture et la qualité des données</u>

Président: Gordon Brackstone
Directeur général,
Direction de la méthodologie

13 h 30 Mesures et ajustements de la couverture
Howard Hogan, chef, Section de la recherche sur le sous-dénombrement, Division de la recherche statistique, Bureau du recensement des États-Unis

Comment rendre l'évaluation de la qualité des données plus pertinente
Richard Burgess, Division des méthodes de recensement et d'enquêtes-ménages

Correction des erreurs non liées à la couverture
Chris Hill, Division des méthodes de recensement et d'enquêtes-ménages

15 h 00 Pause-café

15 h 20 Registres des adresses, avantages et désavantages
David Whitford, chef, Direction de la coordination de la recherche, Division de la planification du recensement décennal, Bureau du recensement des États-Unis

Applications des registres des adresses au recensement canadien
Don Royce, Division des méthodes de recensement et d'enquêtes-ménages

16 h 00 Discussion avec participation de l'assemblée

Le vendredi 11 octobre

9 h 00 - 11 h 50 Séance: <u>Le rôle de la recherche et des essais</u>

Président: E.T. Pryor
Directeur général,
Direction de la statistique démographique et du recensement

9 h 00	Discussion:	La planification des recensements de 1990 – Perspective internationale
	Orateurs:	David Pearce, chef, Division du recensement, Bureau des recensements de la population et des enquêtes démographiques, Royaume-Uni Peter Bounpane, directeur adjoint des recensements démographiques, Bureau du recensement des États-Unis Henry Kriegel, directeur, Planification à long terme et développement des systèmes du recensement de la population, Bureau de la statistique d'Australie
10 h 30	Pause-café	
10 h 50	Discussion:	Maintenant, que faisons-nous?
	Participants:	David Pearce, Peter Bounpane, Henry Kriegel et Edward Pryor
11 h 20	Discussion avec participation de l'assemblée	
11 h 50	**ALLOCUTION DE FERMETURE**	

ANNEXE 2 - LISTE DES PARTICIPANTS

Barnabé, Richard
Directeur régional
Québec
Direction des opérations régionales

Bounpane, Peter A.
Directeur adjoint pour les recensements démographiques
Bureau du recensement des États-Unis

Brackstone, Gordon
Directeur général
Direction de la méthodologie
JT-5B8

Bradley, Ross
Directeur
Division de la géographie
JT-3B7

Burgess, Richard
Chef
Section de la qualité des données et de l'analyse
Division des méthodes de recensement et d'enquêtes-ménages
JT-4C6

Burroughs, Rick
Services aux utilisateurs
Recensement de l'agriculture
Division de l'agriculture et des ressources naturelles
SC-3000

Clayton, Frank
Clayton Research Associates

Code, Oliver G.
Chef
Section des cultures
Division de l'agriculture et des ressources naturelles
SC-2401

Croot, Dave A.
Directeur
Division des services aux clients
SC-2401

Cunningham, Ron
Division des services et du développement informatiques

Deecker, Gordon
Chef
Centre de géocartographie
Division des services et du développement informatiques
JT-2A2

Fellegi, Ivan P.
Statisticien en chef
R.H. Coats 26-A

Gigantes, Terry
Directeur général
Direction de la statistique sur les ressources, la technologie et les services
JT-13B6

Hicks, Doug
Chef
Section de la planification des opérations et de l'élaboration des procédures du recensement
Division des opérations des enquêtes
JT-6C7

Hill, Chris
Chef
Groupe d'étude national des données
R.H. Coats 11-C

Hogan, Howard
Chef
Recherche sur le sous-dénombrement
Division des recherches statistiques
Bureau du recensement des États-Unis

Jackson, Arnold A.
Chef
Division des opérations décennales
Bureau du recensement des États-Unis

Jeays, Mike
Directeur adjoint
Sous-division de la recherche et des systèmes généraux
Division des services et du développement informatiques
R.H. Coats 13-A

Jones, Mel
Chargé du recensement
Recensement de l'agriculture
Division de l'agriculture et des ressources naturelles
SC-3000

Kazmaier, John A. Jr.
Chef adjoint de division
Division des recensements sur le terrain
Bureau du recensement des États-Unis

Kettle, John
Futuresearch Publishing, Inc.

Kidd, Karole
Division des services et du développement informatiques

Kriegel, Henry
Directeur
Recensement de la population
Développement des systèmes et de la planification à long terme
Bureau de la statistique de l'Australie

Lyberg, Lars
Service de recherches statistiques
Statistique Suède

Macredie, Ian
Directeur
Division de l'analyse des enquêtes des ménages et du travail
JT-6A8

March, Mary
Méthodologiste principale
Division des méthodes de recensement et d'enquêtes-ménages
JT-4B6

Meltz, Noah
Université de Toronto

Miskura, Susan
Chef
Planification décennale
Bureau du recensement des États-Unis

Page, Jerry C.
Directeur régional
Alberta, nord de la Saskatchewan et T.N.-O.
Direction des opérations régionales

Parenteau, Robert
Agent
Section de la délimitation spatiale et de l'analyse
Division de la géographie
JT-3B7

Parker, Jean-Pierre
Fichier de base du géocodage
Section des systèmes spatiaux
Division de la géographie
JT-3A6

Pautler, Charles P.
Chef
Division de l'agriculture
Bureau du recensement des États-Unis

Pearce, David
Chef
Division du recensement
Bureau des recensements de la population et des enquêtes
Royaume-Uni

Podehl, Martin
Directeur
Division des services et du développement informatiques
R.H. Coats 13-A

Pryor, Edward T.
Directeur général
Direction de la statistique démographique et du recensement
JT-5B8

Riddle, John
Directeur général associé
Direction des opérations régionales
JT-6C8

Royce, Don
Chef
Section des opérations
Division des méthodes de recensement et d'enquêtes-ménages
JT-4B5

Ryten, Jacob
Statisticien en chef adjoint
Secteur de l'informatique et de la méthodologie
JT-13B8

Tomasi, Silla G.
Chef adjoint de division pour les opérations
Division de la géographie
Bureau du recensement des États-Unis

Underhay, Boyd J.
Directeur régional
Terre-Neuve et Labrador
Direction des opérations régionales

Whitford, David C.
Chef
Direction de la co-ordination de la recherche
Division de la planification décennale
Bureau du recensement des États-Unis

Williams, Brian J.
Directeur adjoint
Bureau régional du Manitoba et du sud de la Saskatchewan
Direction des opérations régionales

Witiuk, Sid
Directeur adjoint
Sous-division de la géocartographie
Division des services et du développement informatiques
JT-2A5

Yan, Joel
Chef
Méthodologie
Sous-division de la géocartographie
Division des services et du développement informatiques
JT-2A2